Introducing Minkowski Normality

Constructing a singular normal number

Master's thesis of M.R. de Lepper

Supervised by

Dr. K. Dajani (UU) and Prof. Dr. E.A. Robinson, Jr. (GWU)
&
Prof. Dr. S.M. Verduyn Lunel (UU)

Mathematical Institute, Utrecht University

Department of Mathematics, The George Washington University

January 20, 2019

Abstract

We introduce the concept of Minkowski normality; a new type of normality that is related to the continued fraction expansion. Moreover, we use the ordering of rationals that is obtained from the Kepler tree to show that the sequence

$$\frac{1}{2}, \quad \frac{1}{3}, \frac{2}{3}, \quad \frac{1}{4}, \frac{3}{4}, \frac{2}{5}, \frac{3}{5}, \quad \frac{1}{5}, \ldots$$

can be used to give a concrete construction of an infinite continued fraction expansion of which the digits are distributed according to the Minkowski question mark measure. We define an explicit correspondence between continued fraction expansions and binary codes to show that we can use the dyadic Champernowne number to prove normality of the constructed number. Furthermore we provide a generalised construction that is based on the underlying structure of the Kepler tree. This generalisation shows that any construction that concatenates the continued fraction expansions of all rationals, ordered increasingly, based on the sum of the digits of their continued fraction expansion, results in a number that is Minkowski normal.

Contents

Chapter 1

Introduction

All numbers are normal Lebesgue almost everywhere. This was proved by Émile Borel in 1909, who introduced the notion of a normal number. Informally, the normality of a number is defined as a distribution property of an infinite number expansion. For example, consider the irrational number π. It is believed that the decimal expansion of π contains as many occurrences of 0s as it contains occurrences of 1s, 2s, ... or 9s. Moreover, it is strongly conjectured that any finite combination of digits occurs as often as any other finite combination of digits of that same length. If a number exhibits such a property, we say that the number is normal in base 10. In other words, normality corresponds to a distribution property of a number expansion. In the case of a decimal expansion, the distribution is uniform and hence originates from the Lebesgue measure. Although Borel proved that there is an abundance of normal numbers, the only known normal numbers are artificial numbers. The difficulty in studying the normality of a number, is that one needs to consider the behaviour of an infinite expansion. The only way to determine the distribution of an infinite sequence of digits, is to recognise a pattern in the expansion. However, most number expansions do not exhibit patterns and, if they exists, they are hard to detect. As for π, we do not know whether or not there is a pattern, but there is a *conjecture* that it is normal in base 10. Hence, one believes that it is normal, but cannot prove it. Fortunately, there exist concrete numbers that are proven to be normal. The most well known example is

$$0.\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14\ \cdots,$$

which is obtained by concatenating all natural numbers. This result is due to David Champernowne [9]. Other examples of normal numbers, include the number that is obtained by concatenating all squares

$$0.\ 1^2\ 2^2\ 3^2\ 4^2\ 5^2\ 6^2\ 7^2\ 8^2\ 9^2\ 10^2\ 11^2\ \cdots,$$

and the number obtained by concatenating all primes

$$0.\ 2\ 3\ 5\ 7\ 11\ 13\ 17\ 19\ 23\ 29\ 31\ 37\ 41\ \cdots.$$

These numbers have in common that they are *constructed*. The underlying idea of such constructions is that, by defining a pattern, one knows the behaviour of the sequence. Therefore, normality can be proved or disproved.

Over the years, many constructions have been done both of normal numbers, as introduced by Borel, as for other types of normality. These different types of normality correspond to different number expansions and different measures. The concrete constructions that have been developed, are all associated to a distribution that results from the Lebesgue measure. However, in this thesis, we consider a measure that is singular with respect to the Lebesgue measure. That is, we consider the *Minkowski question mark measure*. This measure is specified by the following distribution function

$$?(x) := 2\sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{2^{a_1(x)+a_2(x)+\cdots+a_i(x)}},$$

where $a_i(x)$ comes from the *continued fraction expansion* of $x \in [0,1)$, $i \geq 1$. That is, any real number x can be represented as a - possibly finite - continued fraction expansion

$$x = \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{a_3(x) + \cfrac{1}{\ddots}}}},$$

where $a_i(x) \in \mathbb{N}$ for all $i \geq 1$. Using this number expansion and measure, we introduce a different type of normality for the continued fraction expansion. We refer to this type of normality as *Minkowski normality*. Informally, we say that a number x is Minkowski normal if the digits $(a_i(x))_{i\geq 1}$ are distributed according to the Minkowski question mark measure.

The main goal of the thesis is to construct a Minkowski normal number $\mathcal{K}$. As such, we construct an infinite continued fraction expansion and show that the corresponding sequence of digits is distributed according to the Minkowski question mark measure. This forms the main result of the thesis, which is stated in Theorem 4.13. Specifically, for the construction we consider the ordering of rationals that is given by the *Kepler tree*. This is a binary tree that orders the rationals in the unit interval, based on the sum of the digits of their continued fraction expansion. The constructed number is obtained by concatenating the continued fraction expansions of the rationals using the Kepler order. For the proof of normality, we show that there is a correspondence between binary codes and rationals in the Kepler tree. Moreover, we show that we can use binary codes to determine the distribution of the sequence of digits that represent the constructed number.

The thesis has the following structure. In chapter 2 we discuss the mathematics underlying this thesis. That is, we introduce *ergodic theory*, theory on continued fractions and some background on the Minkowski question mark measure. Subsequently, chapter 3 contains a literature study on normal numbers and related results. Here we formally define different notions of normality and discuss results obtained so far. The importance of this chapter lies with the techniques that have been used in the proofs. Chapter 4 focusses on the construction and proof that $\mathcal{K}$ is a Minkowski normal number. Then, we reflect on the construction and show that we can generalise our result. We also use this reflection and generalisation to suggest topics for further research. Finally, we summarise our results in chapter 6 .

CHAPTER 2

Mathematical preliminaries

The main goal of this thesis is to construct an infinite sequence of digits that exhibits some distribution property. In order to study the behaviour of sequences, we first discuss notions and results from *ergodic theory*. In most cases we provide examples of these notions and results directly after introducing them. Sometimes, however, this is not the case and the usefulness of a notion or result becomes apparent in a future section. After we have become familiar with ergodic theory, we introduce a type of number expansion called a *continued fraction expansion*. Then, we introduce the *Gauss map*, which allows us to study the behaviour of the digits of the continued fraction expansion. We conclude the chapter by introducing the *Minkowski question mark function*, which specifies the distribution of the constructed sequence.

2.1.0 *Ergodic Theory*

In this section we describe notions and results from ergodic theory that underlie this thesis. The mathematics involved can be related to numerous fields. One of the central concerns is the behaviour of systems that evolve over time. More specifically, ergodic theory is the study of asymptotic behaviour of averages over space and time. Results from ergodic theory provide conditions under which these averages coincide. As we work with probability spaces, we introduce the notation and results from this perspective. The main goal of this section is to provide the reader with a general framework in ergodic theory that is sufficient to understand the - underlying - mathematics that is used in future sections. The section is based on [11, Chapters 1-5].

Let $(X, \mathcal{F}, \mu)$ denote a probability space, where the space X is formed by the collection of all states of the system. The dynamics of a system are then represented by a measurable map $T : X \to X$, such that for $x \in X$, Tx is the state of the system at time $t = 1$. Furthermore, we define the *orbit* of x as the sequence $(T^i x)_{i\geq 0}$. Then, for $f \in L_1(\mu)$, we define the *time average* as

$$\hat{f}(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x), \tag{2.1}$$

and the *space average* as

$$\bar{f} := \int_X f d\mu. \tag{2.2}$$

From these definitions it can be seen that the orbits play a central role when studying the behaviour of dynamical systems. Ergodic theory is related to the question of when these limits exist (a.e.) and under what conditions they coincide. For both cases, it turns out to be important that the orbit is stationary. In other words, we want the map T to be *measure preserving*.

Definition 2.1 (Measure preserving)**.** [11, Definition 1.2.1] Let $(X, \mathcal{F}, \mu)$ be a probability space, and $T : X \to X$ measurable. The map T is said to be measure preserving with respect to μ if $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{F}$. Furthermore, if T is measure preserving w.r.t. μ, then $(X, \mathcal{F}, \mu, T)$ is called a *measure preserving system.*

Example 2.2 (Bernoulli shift)**.** This example is based on [11, Example 1.3.6].
Let $X = \{0, 1, \ldots, b-1\}^{\mathbb{N}}$ and let $\mathcal{F}_b$ be the σ-algebra on X that is generated by the cylinders. Furthermore, let $p_b = (p_0, p_1, \ldots, p_{b-1})$ be a positive probability vector, which we use to define the *Bernoulli measure* $\mu_b := \{p_0, p_1, \ldots, p_{b-1}\}^{\mathbb{N}}$. Given a cylinder set $A = \{x : x_0 = a_0, x_1 = a_1, \ldots, x_n = a_n\} \in \mathcal{F}_b$, its measure is

$$\mu_b(A) = p_{a_0} p_{a_1} \cdots p_{a_n}.$$

Let $T_b : X \to X$ be the transformation that is defined by $T_b x = y$, where $y_n = x_{n+1}$. We have that $T^{-1}A = \bigcup_{i=0}^{b-1}\{x : x_0 = i, x_1 = a_0, \ldots, x_{n+1} = a_n\} \in \mathcal{F}_b$, which has measure

$$\mu_b(T^{-1}A) = \sum_{i=0}^{b-1} p_i p_{a_0} p_{a_1} \cdots p_{a_n} = p_{a_0} p_{a_1} \cdots p_{a_n} \sum_{i=0}^{b-1} p_i = p_{a_0} p_{a_1} \cdots p_{a_n} = \mu_b(A).$$

We conclude that the *Bernoulli shift* T_b is a measure preserving transformation. This map is also referred to as the *left shift.*

The notions of normality that are discussed in chapter 3 all result from the fact that the distribution property comes from the distribution that is naturally associated to the type of number expansion that is chosen. The chosen expansion is associated to a measure preserving transformation that generates the digits of the corresponding expansion. For a moment, let T denote such a transformation and let $(X, \mathcal{F}, \mu, T)$ be a measure preserving system. The dynamics of this system are represented by T, which moves the points of X. The uncertainty of where T moves a point is also called randomness; *"the quality or state of lacking a pattern or principle of organization; unpredictability"*[1]. One way to quantify the amount of randomness that is generated by T, is by looking at the entropy of a transformation. The entropy in the system $(X, \mathcal{F}, \mu, T)$ varies for different μ and the measure of maximal entropy is the distribution that is naturally associated to a type of expansion.

However, in this thesis we do not consider entropy. Instead, we consider the influence of a transformation T on a random variable Z that is defined on X. This influence is described by the *Perron-Frobenius operator* P_μ of T under μ. This operator defines how the distribution of a random variable Z evolves under iterations of T. Thus suppose that Z admits a density h with respect to μ, i.e.

$$\mathbb{P}(Z \in A) = \int_A h d\mu,$$

then $T \circ Z$ admits density P_μ^h with respect to μ

$$\mathbb{P}(T \circ Z \in A) = \mathbb{P}(Z \in T^{-1}A) = \int_{T^{-1}A} h d\mu = \int_A P_\mu^h d\mu, \qquad A \in \mathcal{B}.$$

In other words, the Perron-Frobenius operator P_μ of T under μ is the bounded linear operator

[1] The definition according to the Oxford dictionary

in L^1_μ that takes $f \in L^1_\mu$ into $P^f_\mu \in L^1_\mu$ with

$$\int_A P^f_\mu d\mu = \int_{T^{-1}A} f d\mu, \qquad A \in \mathcal{B}, \tag{2.3}$$

see [19, Section 2.1]. Defining $\nu(A) := \int_A P^f_\mu d\mu$, $A \in \mathcal{B}$, we find that ν is absolutely continuous with respect to μ. The Radon-Nikodym theorem then states that every measure ν that is absolutely continuous with respect to μ is of this form [32, Chapter 6]. Hence, it follows that the Perron-Frobenius operator of T under μ is a.e. given by the Radon-Nikodym derivative of ν with respect to μ.

We now prove that measure preservingness corresponds to stationarity of the orbit. Moreover, we prove that for a measurable function $f : X \to \mathbb{R}$ and a measure preserving transformation T, the sequence $(f(T^i x))_{i\in\mathbb{N}}$ is stationary.

Proof. Let $r_1, \dots, r_n$ be integers. Then for all $B_1, \dots, B_n \in \mathcal{B}$ and any $k \geq 1$, we have that

$$\begin{aligned}
\mu\big(\{x : f(T^{r_1}x) \in & B_1, \dots, f(T^{r_n}x) \in B_n\}\big) \\
&= \mu\big(\{x : T^{r_1}x \in f^{-1}(B_1), \dots, T^{r_n}x \in f^{-1}(B_n)\}\big) \\
&= \mu\big(T^{-1}\{x : T^{r_1}x \in f^{-1}(B_1), \dots, T^{r_n}x \in f^{-1}(B_n)\}\big) \\
&= \mu\big(\{x : T(T^{r_1}x) \in f^{-1}(B_1), \dots, T(T^{r_n}x) \in f^{-1}(B_n)\}\big) \\
&= \mu\big(\{x : T^{r_1+1}x \in f^{-1}(B_1), \dots, T^{r_n+1}x \in f^{-1}(B_n\})\big),
\end{aligned}$$

where we used the fact that T is measure preserving in line 2. Repeating the above steps k times, shows that

$$\begin{aligned}
\mu\big(\{x : f(T^{r_1}x) \in & B_1, \dots, f(T^{r_n}x) \in B_n\}\big) \\
&= \mu\big(\{x : f(T^{r_1+k}x) \in B_1, \dots, f(T^{r_n+k}x) \in B_n\}\big).
\end{aligned}$$

We conclude that the sequence $(f(T^i x))_{i\in\mathbb{N}}$ is stationary. $\square$

Thus if a transformation is μ-invariant, it follows that the distribution of a point $x \in X$ does not change over time. When studying the behaviour of the system, we are interested in its dynamics, which can be studied by studying the orbits of points $x \in X$. In 1899, Poincaré proved a simple, yet remarkable result about the behaviour of dynamical systems.

Theorem 2.3 (Poincaré Recurrence Theorem)**.** *[11, Theorem 1.4.1] Let $(X, \mathcal{F}, \mu, T)$ be a measure preserving system. If $A \in \mathcal{F}$ such that $\mu(A) > 0$, then almost all points of A return infinitely often to A under iterations of T.*

Thus, if a system starts in a state that has a positive measure, we know that the system will return to that state infinitely often. Hence, for $x \in A$, there exist infinitely many $n_1, n_2, \dots \in \mathbb{N}$ such that $T^{n_i}x \in A$. Moreover, suppose that A is such that $T^{-1}A = A$. Then it follows that the orbit of x stays within A. That is, $(T^i x)_{i\geq 0} \subset A$ for all $x \in A$. Moreover, as $T^{-1}(X\backslash A) = X\backslash A$, we could study a system's behaviour by decomposing it into two parts. When it is not possible to decompose the state space into two invariant subsets of positive measure, we say that the system is *ergodic*.

Definition 2.4 (Ergodicity)**.** [11, Definition 1.6.1] Let $(X, \mathcal{F}, \mu, T)$ be a measure preserving system. The pair (T, μ) is ergodic if for every $A \in \mathcal{F}$ s.t. $T^{-1}A = A$, we have $\mu(A) \in \{0, 1\}$.

Example 2.5 (Ergodicity of Bernoulli systems)**.** Example 2.2 shows that $(X, \mathcal{F}_b, \mu_b, T_b)$ is a measure preserving system. Recall that $\mathcal{F}_b$ is generated by cylinder sets of the form $\{x : x_0 = a_0, x_1 = a_1, \dots, x_n = a_n\}$. Let

$$\mathcal{A}_i := \sigma(x_i, x_{i+1}, x_{i+2}, \dots), \quad \mathcal{A} := \bigcap_{i \geq 1} \mathcal{A}_i$$

and recall that $\mathcal{A}$ is the tail σ-algebra. Given $A = \{x : x_0 = a_0, x_1 = a_1, \dots, x_n = a_n\} \in \mathcal{F}_b$, we have that $T_b^{-i}A \in \mathcal{A}_i$, for all $i \geq 1$. Moreover, for any T_b-invariant set A, we have that $A = T_b^{-i}A \in \mathcal{A}_i$, for all $i \geq 1$ and hence $A \in \mathcal{A}$. As Kolmogorov's zero-one law states that any set in the tail σ-algebra is trivial, we conclude that $\mu_b \in \{0, 1\}$ and hence that (T_b, μ_b) is ergodic. A different and more detailed proof of ergodicity can for instance be found in [11, Example 1.8.2].

In 1931, the American mathematician D.G. Birkhoff proved what is now known as the Pointwise Ergodic theorem. This theorem states that the time average exists a.e. Additionally, he proved that the time average and space average coincide when the system is ergodic. A proof can for instance be found in Kamae and Keane [20].

Theorem 2.6 (The Pointwise Ergodic Theorem)**.** *[11, Theorem 2.1.1]. Let $(X, \mathcal{F}, \mu, T)$ be a measure preserving system. Then, for all $f \in L_1(\mu)$,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n} f(T^i x) = \hat{f}(x) \tag{2.4}$$

exists almost everywhere, is T-invariant and $\int_X f d\mu = \int_X \hat{f} d\mu$. Moreover, if (T, μ) is ergodic, then $\hat{f}$ is constant a.e. and $\hat{f} = \int_X f d\mu$.

We highlight the importance of this theorem at the end of this section. Furthermore, as a consequence of the Pointwise Ergodic Theorem, we get the following characterisation of ergodicity.

Proposition 2.7. Let $(X, \mathcal{F}, \mu, T)$ be a measure preserving probability system, and $\mathcal{S}$ a generating semi-algebra of $\mathcal{F}$. Then (T, μ) is ergodic if and only if for all $A, B \in \mathcal{S}$, it holds that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \mu(A)\mu(B). \tag{2.5}$$

This new characterisation of ergodicity can for instance be used to give an alternative proof of Example 2.5.

Alternative proof of ergodicity of Bernoulli systems. Let A, B be two cylinders that generate $\mathcal{F}_b$. As cylinders depend on a finite number of coordinates, $\exists N$ such that $\forall n \geq N$, $T_b^{-n}A$ and B are independent. Hence for all $n \geq N$, it holds that

$$\mu_b(T_b^{-n}A \cap B) = \mu_b(A)\mu_b(B).$$

By taking limits on both sides, we find that $\lim_{n\to\infty} \mu_b(T_b^{-i}A \cap B) = \mu_b(A)\mu_b(B)$. Hence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu_b(T_b^{-i}A \cap B) = \mu_b(A)\mu_b(B).$$

We conclude from Proposition 2.7 that the Bernoulli shift is ergodic. □

Proposition 2.7 is one out of many characterisations of ergodicity. Different characterisations make it possible to prove ergodicity in different ways. For instance, Example 2.5 gives a direct proof using probability theory and the previous example gives a more dynamical proof using ergodic theory. As different systems may require different approaches when proving or disproving ergodicity, we introduce some tools that may be helpful. Specifically, the following lemma provides a useful tool when proving ergodicity of a system $([0,1), \mathcal{B}, T, \mu)$, for μ equivalent to the Lebesgue measure λ.

Lemma 2.8 (Knopp's lemma). *[11, Lemma 1.8.1]. Let λ denote the Lebesgue measure. If B is a Lebesgue set and $\mathcal{C}$ a class of subintervals of $[0,1)$, satisfying*

(a) every open subinterval of [0,1) is at most a countable union of disjoint elements from $\mathcal{C}$,

(b) $\forall C \in \mathcal{C}, \lambda(B \cap C) \geq \kappa\lambda(C)$, where $\kappa > 0$ is independent of C,

then $\lambda(B) = 1$.

The usefulness of this lemma can be seen from the case where B is such that $T^{-1}B = B$. The lemma is particularly useful for proving ergodicity of systems that have a measure structure that is equivalent to the Lebesgue measure. For more general cases, we introduce the following.

Theorem 2.9. *[11, Theorem 1.6.1] Let $(X, \mathcal{F}, \mu, T)$ be a dynamical system. The following are equivalent.*

(i) (T, μ) is ergodic.

(ii) If $B \in \mathcal{F}$ with $\mu(T^{-1}B \Delta B) = 0$, then $\mu(B) \in \{0, 1\}$.

(iii) If $A \in \mathcal{F}$ with $\mu(A) > 0$, then $\mu(\cup_{i=1}^{\infty} T^{-i}A) = 1$.

(iv) If $A, B \in \mathcal{F}$ with $\mu(A) > 0$ and $\mu(B) > 0$, then there exists $n > 0$ s.t. $\mu(T^{-n}A \cap B) > 0$.

(v) If f is measurable and a.e. T-invariant, then f is a constant a.e.

Besides providing different tools for proving ergodicity, we can interpret these different characterisations to give intuition for ergodicity. Most importantly, ergodicity tells us something about the behaviour of the system. Whereas Poincaré's recurrence theorem tells us that for a set of positive measure A, almost all points in A will revisit A infinitely often under iterations of T, ergodicity tells us that almost all points in X will visit A infinitely often (iii). This is regardless of where the system starts. Similarly, as $x \in T^{-n}A \cap B \implies T^n x \in A$ and $x \in B$, ergodicity implies that points in a set of positive measure eventually visit other sets of positive measure (iv).

Most of the tools so far, are aimed at proving ergodicity of a measure preserving system by only considering that system. However, we can also prove ergodicity by considering different systems and "extending" ergodicity from one system to the other. In the following theorem, we consider different probability measures on the same underlying measurable space. The theorem states two results relating the different measures. The first part of the theorem allows us to extend ergodicity from one system to another. The second part shows that, if both systems are ergodic, the probability measures should either coincide or be mutually singular.

Theorem 2.10. *[11, Theorem 2.1.2] Let $(X, \mathcal{F}, \mu_1, T)$ and $(X, \mathcal{F}, \mu_2, T)$ be measure preserving systems.*

(i) If (T, μ_1) is ergodic and $\mu_2 \ll \mu_1$, then $\mu_1 = \mu_2$.

(ii) If both (T, μ_1) and (T, μ_2) are ergodic, then either $\mu_1 = \mu_2$ or $\mu_1 \perp \mu_2$.

This theorem allows one to compare two different measure preserving systems, where the only difference in structure is the measure structure. It is also possible, however, to compare two - seemingly - completely different systems. For instance, it is often the case that two systems seem to be completely different, but behave in the same way. When this is the case, we say that two systems are isomorphic. Let $(X, \mathcal{F}, \mu, T)$ and $(Y, \mathcal{G}, \nu, S)$ be two measure preserving systems, then an isomorphism is a map $\phi : (X, \mathcal{F}, \mu, T) \to (Y, \mathcal{G}, \nu, S)$ such that the following holds.

(i) ϕ is bijective almost everywhere. That is, there exists null sets $N_X \subset X$ and $N_Y \subset Y$ such that $\phi : X \setminus N_X \to Y \setminus N_Y$ is a bijection.

(ii) ϕ and ϕ^{-1} are measurable.

(iii) ϕ and ϕ^{-1} are measure preserving: $\nu = \mu \circ \phi^{-1}$. That is, $\nu(B) = \mu(\phi^{-1}(B))$ for all $B \in \mathcal{G}$. Analogous conditions should hold for ϕ^{-1}.

(iv) ϕ and ϕ^{-1} preserve the dynamics: $\phi \circ T = S \circ \phi$. This means that ϕ should be such that

$$\begin{array}{ccccccccccc}
x & \xrightarrow{T} & Tx & \xrightarrow{T} & T^2x & \xrightarrow{T} & \cdots & \xrightarrow{T} & T^n x & \xrightarrow{T} & \\
\big\downarrow{\scriptstyle\phi} & & \big\downarrow{\scriptstyle\phi} & & \big\downarrow{\scriptstyle\phi} & & \big\downarrow{\scriptstyle\phi} & & \big\downarrow{\scriptstyle\phi} & & . \\
\phi(x) & \xrightarrow{S} & S(\phi(x)) & \xrightarrow{S} & S^2\phi(x) & \xrightarrow{S} & \cdots & \xrightarrow{S} & S^n\phi(x) & \xrightarrow{S} &
\end{array}$$

Definition 2.11 (Isomorphism). [11, Definition 3.1.1] Two dynamical systems $(X, \mathcal{F}, \mu, T)$ and $(Y, \mathcal{G}, \nu, S)$ are isomorphic if there exist measurable sets $N \subset X$ and $M \subset Y$ with:

- $\mu(X \setminus N) = \nu(Y \setminus M) = 0$ and
- $T(N) \subset N, S(M) \subset M$,

for which there exists a measurable map $\phi : N \to M$ such that (i)–(iv) are satisfied.

Loosely speaking, ϕ is a measure preserving map such that the following diagram commutes

$$\begin{array}{ccc} (X,\mu) & \xrightarrow{T} & (X,\mu) \\ \big\downarrow \phi & & \big\downarrow \phi \\ (Y,\nu) & \xrightarrow{S} & (Y,\nu) \end{array} .$$

As isomorphisms preserve dynamics, it follows that we can extend ergodicity from one system to another by showing that they are isomorphic. In particular, we have the following.

Proposition 2.12. Let $(X, \mathcal{F}_b, \mu_b, T_b)$ be a Bernoulli system and let $(Y, \mathcal{G}, \nu, S)$ be a measure preserving system. If $(X, \mathcal{F}_b, \mu_b, T_b)$ and $(Y, \mathcal{G}, \nu, S)$ are isomorphic, then (S, ν) is ergodic.

This concludes the general notions and results on ergodic theory. Before turning to the next section, we highlight the importance of Theorem 2.6. The importance of the Pointwise Ergodic Theorem can be illustrated as follows. Suppose that we have an ergodic probability system $(X, \mathcal{F}, T, \mu)$ and we consider the orbit of an arbitrary $x \in X$. More precisely, we consider the asymptotic frequency of visits of x to a set $A \in \mathcal{F}$ under iterations of T. We can represent this asymptotic frequency by the time average of the function $f = \mathbb{1}_A \in L^1_\mu$. Then, by the Pointwise Ergodic Theorem, we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_A(T^i x) = \hat{f}(x) = \int_X \mathbb{1}_A d\mu = \mu(A). \tag{2.6}$$

The second equality results from the fact that the system is ergodic and therefore constant a.e. We conclude that the asymptotic frequency of visits to the set A is precisely the probability of being in A. As the normality of a number is a distribution property of the sequence of digits that represent this number, the previously described illustration will prove to be useful for proving this property. To see this, think of A as a set of states that satisfies some condition. Subsequently, if we can define a transformation T that runs through the sequence of digits that represent x, the left hand side of (2.6) is the asymptotic frequency of digits that satisfy this property. Though it is practically impossible to evaluate this asymptotic frequency, for ergodic systems we know that it is precisely the probability of the system being in a state that satisfies this condition. The next section starts by introducing *continued fractions*, which provides us with a sequence representation of a number. Subsequently, we introduce the transformation that runs through this sequence of digits and study it under two different measures, of which one is invariant and one is not.

2.2.0 *Continued fraction expansions*

Similar as to expressing a number in binary or decimal form, the continued fraction expansion is just another way to represent a number. They are closely related to the Euclidian algorithm and have some useful properties. Continued fractions can for instance be used to find "best" rational approximations of irrational numbers. These and other relevant results on continued fractions are explained in this section. The results presented in this section can be found in [19, Chapter 1], [27, Chapter 5] or any other handbook on continued fractions. From now on, let $X = [0, 1)$ and let $\mathcal{B}$ denote the Borel σ-algebra on X.

Any real number $x \in [0, 1)$, can be represented by its continued fraction expansion

$$x = \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{a_3(x) + \cfrac{1}{\ddots}}}}, \tag{2.7}$$

where the *partial quotiens* $a_i(x)$ are integers for all $i \in \mathbb{N}$. It is common to drop the dependence on x and use the notation $x = [a_1, a_2, a_3, \cdots]$. If x is rational, the continued fraction expansion is finite and non-unique. Non-uniqueness can be seen from the identity $x = [a_1, a_2, \cdots, a_n] = [a_1, a_2, \cdots, a_n - 1, 1]$. This result is summarised below.

Theorem 2.13. *[27, Theorem 5.2] Any finite continued fraction represents a rational number. Conversely, any rational number x can be expanded in a finite continued fraction in exactly two ways.*

On the other hand, the continued fraction expansion of an irrational x is infinite and unique. When we truncate the continued fraction expansion of $x = [a_1, a_2, a_3, \cdots]$ at the n-th digit, we get a rational approximation that we call the *n-th convergent*

$$\omega_n = \frac{p_n}{q_n} = [a_1, a_2, \cdots, a_n], \tag{2.8}$$

where $p_n, q_n \in \mathbb{N}$ and $\gcd(p_n, q_n) = 1$. The integers p_n and q_n are called the *continuants* of x and satisfy the equations

$$p_n = a_n p_{n-1} + p_{n-2}, \tag{2.9}$$

$$q_n = a_n q_{n-1} + q_{n-2}, \tag{2.10}$$

with $p_{-1} = q_0 = 1$ and $p_0 = q_{-1} = 0$. The sequence $(q_i)_{i \geq 1}$ is monotonically increasing and can be used to bound the error in the approximation

$$\frac{1}{2q_n q_{n-1}} < |x - \frac{p_n}{q_n}| < \frac{1}{q_n q_{n+1}}. \tag{2.11}$$

One of the properties of the continued fraction is that the n-th convergent is the best rational approximation of x. That is to say that there is no rational approximation with denominator smaller or equal to q_n that is closer to x. Furthermore, apart from bounding the approximation, we find that the sequence $(q_i)_{i\geq 1}$ also bounds the distance between two consecutive approximations. We provide a short derivation below.

$$\begin{aligned}
\omega_n - \omega_{n-1} &= \frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} \\
&= \frac{p_n q_{n-1} - p_{n-1} q_n}{q_n q_{n-1}} \\
&= \frac{p_n q_{n-1} - p_{n-1}(a_n q_{n-1} + q_{n-2})}{q_n q_{n-1}} \\
&= \frac{q_{n-1}(p_n - a_n p_{n-1}) - p_{n-1} q_{n-2}}{q_n q_{n-1}} \\
&= \frac{-(p_{n-1} q_{n-2} - p_{n-2} q_{n-1})}{q_n q_{n-1}} \\
&= \dots \\
&= \frac{(-1)^n (p_0 q_{-1} - p_{-1} q_0)}{q_n q_{n-1}} = \frac{(-1)^{n+1}}{q_n q_{n-1}}.
\end{aligned}$$

Subsequently, note that

$$\omega_n = \sum_{i=1}^{n} \omega_i - \omega_{i-1} = \sum_{i=1}^{n} \frac{(-1)^{i+1}}{q_i q_{i-1}}.$$

As $(q_n)_{n\geq 1}$ is monotonically increasing, the sequence $(\frac{1}{q_i q_{i-1}})_{i\geq 1}$ is monotonically decreasing. From Leibniz' theorem, it follows that $\lim_{n\to\infty} \omega_n$ exists. As we will be constructing a normal number in chapter 4, this is an important result. In particular, because it implies the following: when we are given a sequence of integers $(a'_i)_{i\geq 1} \subset \mathbb{N}$ and define ω'_n as the rationals obtained by forming the finite continued fraction $[a'_1, a'_2, \cdots, a'_n]$, then the sequence $(\omega'_i)_{i\geq 1}$ converges. Moreover, it converges to a unique irrational. This leads to the following proposition.

Proposition 2.14. [11, Proposition 4.1.1] Let $(a_i)_{i\geq 1}$ be a sequence of positive integers, and define the sequence of rationals $(\omega_i)_{i\geq 1}$ as

$$\omega_n := [a_1, a_2, \cdots, a_n], \qquad n \geq 1.$$

Then there exists a unique irrational number $x \in (0, 1)$ such that

$$\lim_{n\to\infty} \omega_n = x.$$

Moreover, we have that $x = [a_1, a_2, a_3, \cdots]$.

Summarising the above, we get the following.

Theorem 2.15. *[27, Theorem 5.11] Every irrational number $x \in [0, 1)$ has a unique representation as an infinite continued fraction $[a_1, a_2, a_3, \cdots]$ and conversely. The integers a_i are positive* *for $i \geq 1$. The n-th convergent $\omega_n = \frac{p_n}{q_n}$, is the finite continued fraction $[a_1, a_2, \cdots, a_n]$. The denominators q_i form a monotonically increasing sequence, for $i \geq 1$. The even and odd convergents are monotonically increasing, respectively decreasing, with x as a limit*

$$0 = \omega_0 < \omega_2 < \omega_4 < \cdots x \cdots < \omega_3 < \omega_1 < 1.$$

Thus, opposite to irrationals, rationals have a finite continued fraction expansion. From this point onwards, we use the convention that any continued fraction $[a_1, a_2, \cdots, a_n]$ is written in its reduced form. That is, we use the convention that $a_n \geq 2$, in which case rationals also have "unique" continued fraction expansions. We next introduce the measurable map that generates the digits of the continued fraction expansion.

2.3.0 *The Gauss map*

In the previous section, we have provided basics results on continued fractions. From Proposition 2.14, we see that an infinite sequence of digits $(a_i)_{i \geq 1}$ can be used to construct a unique irrational number. This is important because it ensures that the number that is constructed in chapter 4 is indeed a unique irrational. In order to study the digits of the continued fraction expansion, we introduce a measure preserving map that represents the dynamics of such a number system.

Continued fractions are closely related to the Euclidian algorithm. One reason for this, is that the digits of the continued fraction expansion can be generated through this algorithm. However, rather than using a number theoretic approach, we use one from ergodic theory to obtain the digits. As such, we introduce the *Gauss map*, show that it generates the digits of the continued fraction expansion and discuss related results.

Definition 2.16 (Gauss map). [11, Example 1.5.3] The Gauss map is the measurable map $\mathcal{T} : X \to X$ with

$$\mathcal{T}x = \frac{1}{x} \mod 1 = \begin{cases} \frac{1}{x} - \lfloor \frac{1}{x} \rfloor & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases} \tag{2.12}$$

When we denote x by its - possibly finite - continued fraction expansion $[a_1, a_2, a_3, \cdots]$ and rewrite equation (2.12) for $x \neq 0$, we find that

$$\begin{aligned} \mathcal{T}x &= \frac{1}{[a_1, a_2, a_3, \cdots]} - \left\lfloor \frac{1}{[a_1, a_2, a_3, \cdots]} \right\rfloor = a_1 + [a_2, a_3, a_4, \cdots] - \left\lfloor a_1 + [a_2, a_3, a_4, \cdots] \right\rfloor \\ &= [a_2, a_3, a_4, \cdots]. \end{aligned}$$

In a similar fashion, we find that the n-th element of the orbit of x is given by

$$\mathcal{T}^{n-1}x = [a_n, a_{n+1}, a_{n+2}, \cdots]. \tag{2.13}$$

When we invert this equation we see that

$$\begin{aligned} \frac{1}{\mathcal{T}^{n-1}x} &= \frac{1}{[a_n, a_{n+1}, a_{n+2}, \cdots]} \\ &= a_n + [a_{n+1}, a_{n+2}, a_{n+3}, \cdots], \end{aligned}$$

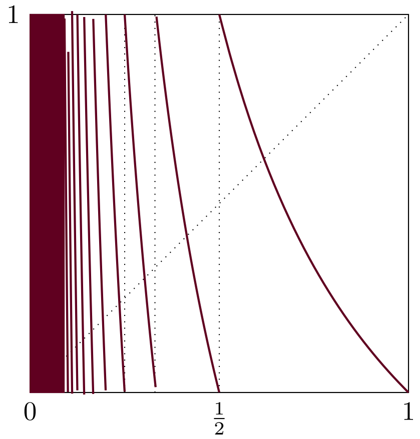


Figure 2.1: The Gauss map $\mathcal{T}$.

from which it follows that the digits a_n are given by $a_n = \lfloor \frac{1}{\mathcal{T}^{n-1}x} \rfloor$. That is, the partial quotients are generated by the Gauss map. Lastly, it is useful to notice that due to (2.13), T is called a *left shift* for the continued fraction and we have that

$$x = [a_1, a_2, \cdots, a_n + \mathcal{T}^n x]. \tag{2.14}$$

As we will be looking at the distribution of the sequence $(a_i)_{i\geq 1}$, we consider them to be $\mathbb{N}$-valued random variables defined on the probability space $(X, \mathcal{B}, \mu)$. When studying these random variables, we prefer a transformation T to be measure preserving with respect to the probability measure μ. This is due to the fact that measure preservingness ensures stationarity of the sequence $(f(T^i x))_{i\geq 1}$, for all measurable f. Moreover, if we can find a measure μ such that (T, μ) is ergodic, Theorem 2.6 ensures that time and space averages coincide. When we consider the various probability measures, it is natural to first look at the Lebesgue measure. However, it turns out that the Gauss map $\mathcal{T}$ is not measure preserving with respect to the Lebesgue measure λ. In order to show this, we first introduce *fundamental intervals.*

Definition 2.17 (Fundamental intervals)**.** A fundamental interval of order n is the set of numbers in $[0, 1)$ that have the same first n digits in their continued fraction expansion. A fundamental interval of order n is thus defined by

$$\Delta(a_1, a_2, \cdots, a_n) := \{x \in [0, 1) : x = [a_1, a_2, \cdots, a_n, \cdots]\},$$

$a_1, a_2, \cdots, a_n \in \mathbb{N}$.

In order to see that this is indeed an interval, notice that for any $x \in \Delta(a_1, a_2, \cdots, a_n)$ we have that $x = [a_1, a_2, \cdots, a_n + \mathcal{T}^n x]$, see (2.14). As $\mathcal{T}$ is defined on $[0, 1)$, it follows that the interval is defined by the endpoints $[a_1, a_2, \cdots, a_n]$ and $[a_1, a_2, \cdots, a_n + 1]$. It is possible to rewrite the endpoints using the convergents. Doing so leads to

$$\Delta(a_1, a_2, \cdots, a_n) = \begin{cases} [\frac{p_n}{q_n}, \frac{p_n + p_{n-1}}{q_n + q_{n-1}}) & \text{if } n \text{ is even,} \\ [\frac{p_n + p_{n-1}}{q_n + q_{n-1}}, \frac{p_n}{q_n}) & \text{if } n \text{ is odd.} \end{cases} \tag{2.15}$$

A more rigorous proof of the above can for instance be found in [19, Theorem 1.2.2]. Next, we show that the Gauss map is not measure preserving with respect to the Lebesgue measure.

Example 2.18. Consider the fundamental interval of order one $\Delta(1)$. Then using (2.15) and the fact that $p_{-1} = q_0 = 1$ and $p_0 = q_{-1} = 0$, it follows that

$$\Delta(1) = \{x \in [0, 1) : x = [1, \cdots]\} = [\frac{1}{2}, 1).$$

Hence $\lambda(\Delta(1)) = \frac{1}{2}$. As $\mathcal{T}$ is a left shift, we find that $\mathcal{T}^{-1}(\Delta(1)) = \bigcup_{i\geq 1} \Delta(i, 1) = \bigcup_{i\geq 1}(\frac{1}{i+1}, \frac{1}{i+\frac{1}{2}})$. Taking the Lebesgue measure of this set, we see that

$$\begin{aligned} \lambda(\mathcal{T}^{-1}\Delta(1)) &= \lambda\big(\cup_{i\geq 1}(\frac{1}{i+1}, \frac{1}{i+\frac{1}{2}})\big) \\ &= \sum_{i\geq 1} \lambda\big((\frac{1}{i+1}, \frac{1}{i+\frac{1}{2}})\big) \\ &= \sum_{i\geq 1} \frac{1}{i+\frac{1}{2}} - \frac{1}{i+1} \\ &= 2\sum_{i\geq 1} \frac{1}{2i+1} - \frac{1}{2i+2} \\ &= 2\log(2) - 1 \neq \frac{1}{2} = \lambda(\Delta(1)). \end{aligned}$$

Hence we conclude that $\mathcal{T}$ is not measure preserving w.r.t. λ.

The above (counter)example implies that $(X, \mathcal{B}, \lambda, \mathcal{T})$ is not a measure preserving system. Therefore, the system is not ergodic. However, one of the measures that does satisfy our wishes concerning measure preservingness and ergodicity, is the *Gauss measure*. Though it remains unknown how he came to the conclusion, it was Gauss who stated that the density of the partial quotients (with respect to λ) is given by

$$\frac{1}{\log 2} \cdot \frac{1}{1+x},$$

which leads to the definition of the *Gauss measure* γ

$$\gamma(A) := \frac{1}{\log 2} \int_A \frac{1}{1+x} dx \qquad A \in \mathcal{B}. \tag{2.16}$$

As $x \in [0,1)$, we find that

$$\frac{1}{2\log 2}\lambda(A) \leq \gamma(A) \leq \frac{1}{\log 2}\lambda(A). \tag{2.17}$$

Hence, we have that $\gamma \ll \lambda$ (and $\lambda \ll \gamma$). It follows that the Perron-Frobenius operator of the Gauss map under the Lebesgue measure is given by the Radon-Nikodym derivative of the Gauss measure with respect to the Lebesgue measure. In order to show that, opposite to the Lebesgue measure, the Gauss measure is preserved by the Gauss map, we note that it is sufficient to prove invariance of an interval $(a,b) \subset [0,1)$. This is due to the fact that the Borel σ-algebra is generated by such intervals. We find that

$$\mathcal{T}^{-1}(a,b) = \{x \in [0,1) : \mathcal{T}x = \frac{1}{x} - a_1(x) \in (a,b)\},$$

where $a_1(x) \in \mathbb{N}$. Therefore

$$\mathcal{T}^{-1}(a,b) = \bigcup_{i\geq 1}(\frac{1}{i+b}, \frac{1}{i+a}), \tag{2.18}$$

and we find that

$$\begin{aligned}
\gamma(\mathcal{T}^{-1}(a,b)) &= \gamma\big(\cup_{i\geq 1}(\frac{1}{i+b}, \frac{1}{i+a})\big) \\
&= \sum_{i\geq 1}\gamma\big((\frac{1}{i+b}, \frac{1}{i+a})\big) \\
&= \sum_{i\geq 1}\frac{1}{\log 2}\int_{\frac{1}{i+b}}^{\frac{1}{i+a}}\frac{1}{1+x}dx \\
&= \frac{1}{\log 2}\sum_{i\geq 1}\log(1+\frac{1}{i+a}) - \log(1+\frac{1}{i+b}) \\
&= \frac{1}{\log 2}\sum_{i\geq 1}\log\big(\frac{i+a+1}{i+a}/\frac{i+b+1}{i+b}\big) \\
&= \frac{1}{\log 2}\lim_{n\to\infty}\sum_{i=1}^{n}\log(i+a+1) - \log(i+a) + \log(i+b) - \log(i+b+1) \\
&= \frac{1}{\log 2}\lim_{n\to\infty}\log(i+a+1) - \log(1+a) + \log(1+b) - \log(i+b+1) \\
&= \frac{1}{\log 2}\log(\frac{1+b}{1+a}) \\
&= \frac{1}{\log 2}\big(\log(1+b) - \log(1+a)\big) \\
&= \frac{1}{\log 2}\int_a^b\frac{1}{1+x}dx
\end{aligned}$$

$$= \gamma\big((a,b)\big).$$

We conclude that the Gauss map is measure preserving with respect to the Gauss measure. Moreover, the Gauss map is ergodic under the Gauss measure. We show this through an application of Knopp's Lemma.

Theorem 2.19. *[11, Theorem 4.2.1] The pair $(\mathcal{T}, \gamma)$ is ergodic.*

Proof. Note that equivalence of the Gauss and Lebesgue measure follows from (2.17). Furthermore, let $[a, b)$ be a subinterval of $[0, 1)$ and $\Delta_n = \Delta(a_1, a_2, \cdots, a_n)$ be a fundamental interval of order n. Then, analogous to (2.15), one can show that $\mathcal{T}^{-n}(a, b) \cap \Delta_n$ is the interval that is specified by the endpoints $[a_1, a_2, \cdots, a_n + a]$ and $[a_1, a_2, \cdots, a_n + b]$

$$\mathcal{T}^{-n}(a,b) \cap \Delta_n = \begin{cases} \left(\dfrac{p_{n-1}a + p_n}{q_{n-1}a + q_n}, \dfrac{p_{n-1}b + p_n}{q_{n-1}b + q_n}\right) & \text{if } n \text{ is even,} \\ \left(\dfrac{p_{n-1}b + p_n}{q_{n-1}b + q_n}, \dfrac{p_{n-1}a + p_n}{q_{n-1}a + q_n}\right) & \text{if } n \text{ is odd.} \end{cases}$$

Without loss of generality, let n be even. The Lebesgue measure of this interval can then be rewritten as

$$\begin{aligned} \lambda(\mathcal{T}^{-n}(a,b) \cap \Delta_n) &= \frac{p_{n-1}b + p_n}{q_{n-1}b + q_n} - \frac{p_{n-1}a + p_n}{q_{n-1}a + q_n} \\ &= \frac{(p_{n-1}b + p_n)(q_{n-1}a + q_n) - (p_{n-1}a + p_n)(q_{n-1}b + q_n)}{(q_{n-1}b + q_n)(q_{n-1}a + q_n)} \\ &= (b-a)\frac{p_{n-1}q_n - p_n q_{n-1}}{(q_{n-1}b + q_n)(q_{n-1}a + q_n)} \\ &= (b-a)\frac{p_{n-1}q_n - p_n q_{n-1}}{q_n(q_n + q_{n-1})} \frac{q_n(q_n + q_{n-1})}{(q_{n-1}b + q_n)(q_{n-1}a + q_n)} \\ &= \lambda(a,b)\lambda(\Delta_n)\frac{q_n(q_n + q_{n-1})}{(q_{n-1}b + q_n)(q_{n-1}a + q_n)}. \end{aligned}$$

Using the fact that $0 \leq a < b < 1$ and that $(q_i)_{i \geq 0}$ is monotonically increasing, we can bound the latter fraction from both above and below to see that

$$\frac{1}{2} < \frac{q_n(q_n + q_{n-1})}{(q_{n-1}b + q_n)(q_{n-1}a + q_n)} < 2.$$

Combining this bound with the one in (2.17), we find that

$$\gamma(\mathcal{T}^{-n}(a,b) \cap \Delta_n) \geq \frac{\log 2}{4}\gamma(\mathcal{T}^{-n}(a,b))\gamma(\Delta_n).$$

Moreover, for B such that $T^{-1}B = B$ and $\gamma(B) > 0$, it holds that

$$\gamma(B \cap \Delta_n) \geq \frac{\log 2}{4}\gamma(B)\gamma(\Delta_n).$$

Therefore, the desired result is obtained by an application of Knopps lemma with $\mathcal{C}$ the collection of all fundamental intervals Δ_n and $\kappa = \frac{\log 2}{4}\gamma(B)$. $\square$

The main goal of this thesis is to construct a Minkowski normal number. As such, the last part of the mathematical framework discusses the distribution that is associated with this new type of normality: the *Minkowski question mark function.*

2.4.0 *Minkowski's $?(\cdot)$ function*

The Minkowski $?(\cdot)$ function is a strictly monotone, continuous and singular function. The function was first introduced in 1904 by Herman Minkowski, whose motivation was to illustrate the following condition for quadratic irrationals [18]

> "*A real number is a quadratic irrational if and only if its continued fraction expansion is infinite and periodic; it is a rational if and only if its continued fraction expansion is finite.* Minkowski's function leads to the following criterion: *x is a quadratic irrational if and only if $?(x)$ is a non-dyadic rational; x is rational if and only if $?(x)$ is a dyadic rational.*"

The question mark function can be constructed in several ways. We next present the basis of the construction as in Salem [30], after which we formally define the Minkowski question mark measure and discuss some of its properties and related results. A generalised form of the construction can be found in [18].

We construct the Minkowski question mark function $?(\cdot)$ by defining a sequence of sets $(\mathcal{M}_i)_{i\geq 0}$ and the corresponding values of $?(x)$ for all $x \in \mathcal{M}_n$, $n \geq 0$. Let $\mathcal{M}_0 = \{\frac{0}{1}, \frac{1}{1}\}$ and define the base condition

$$?(\tfrac{0}{1}) = 0 \qquad \text{and} \qquad ?(\tfrac{1}{1}) = 1.$$

At the next step, we take the mediant of $\frac{0}{1}$ and $\frac{1}{1}$ and add this to $\mathcal{M}_0$ to form $\mathcal{M}_1 = \{\frac{0}{1}, \frac{1}{2}, \frac{1}{1}\}$. Subsequently, we define

$$?(\tfrac{0+1}{1+1}) = ?(\tfrac{1}{2}) = \tfrac{1}{2}\big(?(\tfrac{0}{1}) + ?(\tfrac{1}{1})\big) = \tfrac{1}{2}.$$

This results in the first order approximation of the function. By repeating process, we define the Minkowski question mark function for all $x \in [0, 1]$. That is, $\mathcal{M}_n$ is formed from $\mathcal{M}_{n-1}$ by inserting the mediant between two adjacent fractions in $\mathcal{M}_{n-1}$. The corresponding function value is then given by the arithmetic mean of the function values of the fractions that make up the mediant. Let $\frac{p}{q}$ and $\frac{r}{s}$ be two adjacent fractions in $\mathcal{M}_{n-1}$ such that $\frac{p}{q} < \frac{r}{s}$. Then we insert the mediant in $\mathcal{M}_n$, for which we have that

$$\frac{p}{q} < \frac{p+r}{q+s} < \frac{r}{s}.$$

Subsequently, we define the corresponding question mark function value by

$$?(\frac{p+r}{q+s}) = \frac{1}{2}\big(?(\frac{p}{q}) + ?(\frac{r}{s})\big). \tag{2.19}$$

The first four sets of $(\mathcal{M}_i)_{i\geq 0}$ are

$$\begin{aligned}
\mathcal{M}_0 &= \{\tfrac{0}{1}, && && && && && && && && \tfrac{1}{1}\},\\
\mathcal{M}_1 &= \{\tfrac{0}{1}, && && && && \tfrac{1}{2}, && && && && \tfrac{1}{1}\},\\
\mathcal{M}_2 &= \{\tfrac{0}{1}, && && \tfrac{1}{3}, && && \tfrac{1}{2}, && && \tfrac{2}{3}, && && \tfrac{1}{1}\},\\
\mathcal{M}_3 &= \{\tfrac{0}{1}, && \tfrac{1}{4}, && \tfrac{1}{3}, && \tfrac{2}{5} && \tfrac{1}{2}, && \tfrac{3}{5} && \tfrac{2}{3}, && \tfrac{3}{4}, && \tfrac{1}{1}\}.
\end{aligned}$$

Using equation (2.19), we find that the nth-order approximation of the question mark function maps $\mathcal{M}_n$ to the set of dyadic rationals of order n; $\mathcal{D}_n = \{\frac{k}{2^n} : k = 0, 1, \ldots, 2^n\}$. The first four levels of $(\mathcal{D}_i)_{i\geq 0}$ are given below

$$\begin{aligned}
\mathcal{D}_0 &= \{\tfrac{0}{2^0}, && && && && && && && && \tfrac{1}{2^0}\},\\
\mathcal{D}_1 &= \{\tfrac{0}{2^1}, && && && && \tfrac{1}{2^1}, && && && && \tfrac{2}{2^1}\},\\
\mathcal{D}_2 &= \{\tfrac{0}{2^2}, && && \tfrac{1}{2^2}, && && \tfrac{2}{2^2}, && && \tfrac{3}{2^2}, && && \tfrac{4}{2^2}\},\\
\mathcal{D}_3 &= \{\tfrac{0}{2^3}, && \tfrac{1}{2^3}, && \tfrac{2}{2^3}, && \tfrac{3}{2^3} && \tfrac{4}{2^3}, && \tfrac{5}{2^3} && \tfrac{6}{2^3}, && \tfrac{7}{2^3}, && \tfrac{8}{2^3}\}.
\end{aligned}$$

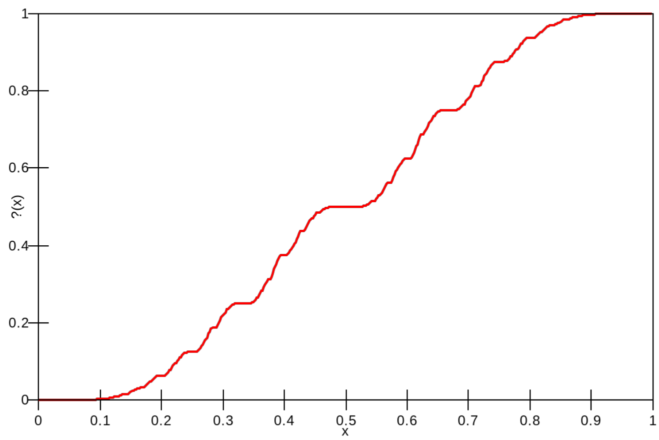


Figure 2.2: The Minkowski question mark.

The figures in Appendix B show the approximations of orders $n = 1, 2, 3$ and 4. Figure 2.2 displays the Minkowski question mark function, which is obtained by the limit of the previously described process. Note that the construction defines the Minkowski question mark function for rationals. However, it follows from continuity that the function is defined for every $x \in [0, 1]$ [30].

The first to provide a detailed study of the function, was Arnoud Denjoy [13]. Among other things, Denjoy expressed the function as a summation and proved its singularity. We use Denjoy's summation to formally define Minkowski question mark measure, after which we discuss some properties and related results.

Definition 2.20 (Minkowski's question mark measure)**.** The Minkowski question mark measure $\mu_?$ is given by the distribution function $?(\cdot)$ that is defined by

$$?(x) := \mu_?((0,x]) = \begin{cases} 2\sum_{i=1}^{n} \frac{(-1)^{i+1}}{2^{a_1+a_2+\cdots+a_i}} & \text{if } x = [a_1, a_2, \cdots, a_n]. \\ 2\sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{2^{a_1+a_2+\cdots+a_i}} & \text{if } x = [a_1, a_2, a_3 \cdots], \end{cases} \tag{2.20}$$

with $x \in [0, 1)$. Furthermore $?(0) = 0$ and $?(1) = 1$.

Note that for any fundamental interval $\Delta_n = \Delta(a_1, a_2, \cdots, a_n)$, $\mu_?(\Delta_n) = 2^{-(a_1+a_2+\cdots+a_n)}$.

It follows from the construction that the Minkowski question mark function is continuous and strictly monotone, which implies that it is differentiable almost everywhere. Although the derivative exists almost everywhere, it is 0 a.e. In turn, this implies that the measure is singular with respect to the Lebesgue measure. Moreover, the Radon Nikodym derivative does not exist and there is no density function in the classical sense. However, Vepstas [36] provides an explicit construction for the derivative of $?(x)$, expressing it as an infinite product of piecewise continuous functions. By doing so, they provide insight into the sets for which the derivative is zero. They show that it vanishes for all rationals and is infinite on the irrationals except on a certain class of quadratic irrationals. A more precise description of these cases can be found in [15].

The fact that the Minkowski question mark measure is singular with respect to the Lebesgue measure can be proved in different ways. A direct proof of singularity can be found in [13] and [30]. In what follows, we show that singularity follows from Theorem 2.10. We first prove that the Gauss map is measure preserving with respect to the Minkowski question mark measure. Subsequently, we prove that the pair $(\mathcal{T}, \mu_?)$ is ergodic and then show that singularity follows from an application of Theorem 2.10(ii).

In order to prove that $\mathcal{T}$ is measure preserving with respect to $\mu_?$, we prove that the measure of an arbitrary interval is preserved under $\mathcal{T}$. For convenience, without loss of generality, let $(a, b) \in [0, 1)$ be an arbitrary interval with rational endpoints. In other words, let (a, b) be an arbitrary interval that has endpoints that are defined by a finite continued fraction. Suppose that $a = [a_1, a_2, \cdots, a_n]$ and $b = [b_1, b_2, \cdots, b_m]$. We then have

$$\begin{aligned} \mu_?\big((a,b)\big) &= ?(b) - ?(a) \\ &= 2\sum_{k=1}^{m} \frac{(-1)^{k+1}}{2^{b_1+b_2+\cdots+b_k}} - 2\sum_{l=1}^{n} \frac{(-1)^{l+1}}{2^{a_1+a_2+\cdots+a_l}}. \end{aligned}$$

Using the preimage of the Gauss map, see (2.18), we find

$$
\begin{aligned}
\mu_?\big(\mathcal{T}^{-1}(a,b)\big) &= \mu_?\big(\cup_{i\geq 1}\,(\frac{1}{i+b},\frac{1}{i+a})\big)\\
&= \sum_{i\geq 1}\mu_?\big((\frac{1}{i+b},\frac{1}{i+a})\big)\\
&= \sum_{i\geq 1} ?\big(\frac{1}{i+a}\big) - ?\big(\frac{1}{i+b}\big)\\
&= \sum_{i\geq 1} ?\big([i,a_1,a_2,\cdots,a_n]\big) - ?\big([i,b_1,b_2,\cdots,b_m]\big)\\
&= \sum_{i\geq 1}\Big(2\big(\frac{1}{2^i}+\frac{1}{2^i}\sum_{l=2}^{n+1}\frac{(-1)^{l+1}}{2^{a_1+a_2+\cdots+a_{l-1}}}\big) - 2\big(\frac{1}{2^i}+\frac{1}{2^i}\sum_{k=2}^{m+1}\frac{(-1)^{k+1}}{2^{b_1+b_2+\cdots+b_{k-1}}}\big)\Big)\\
&= 2\sum_{i\geq 1}\frac{1}{2^i}\big(\sum_{l=2}^{n+1}\frac{(-1)^{l+1}}{2^{a_1+a_2+\cdots+a_{l-1}}} - \sum_{k=2}^{m+1}\frac{(-1)^{k+1}}{2^{b_1+b_2+\cdots+b_{k-1}}}\big)\\
&= 2\sum_{i\geq 1}\frac{1}{2^i}\big(\sum_{k=2}^{m+1}\frac{(-1)^{k}}{2^{b_1+b_2+\cdots+b_{k-1}}} - \sum_{l=2}^{n+1}\frac{(-1)^{l}}{2^{a_1+a_2+\cdots+a_{l-1}}}\big)\\
&= 2\sum_{i\geq 1}\frac{1}{2^i}\big(\sum_{k=1}^{m}\frac{(-1)^{k+1}}{2^{b_1+b_2+\cdots+b_{k}}} - \sum_{l=1}^{n}\frac{(-1)^{l+1}}{2^{a_1+a_2+\cdots+a_{l}}}\big)\\
&= \sum_{i\geq 1}\frac{1}{2^i}\big(2\sum_{k=1}^{m}\frac{(-1)^{k+1}}{2^{b_1+b_2+\cdots+b_{k}}} - 2\sum_{l=1}^{n}\frac{(-1)^{l+1}}{2^{a_1+a_2+\cdots+a_{l}}}\big)\\
&= \sum_{i\geq 1}\frac{1}{2^i}\big(?(b) - ?(a)\big)\\
&= \big(?(b) - ?(a)\big)\sum_{i=1}^{\infty}\frac{1}{2^i}\\
&= \mu_?\big((a,b)\big).
\end{aligned}
$$

In the last step we used the fact that the sum of the geometric series $(2^{-i})_{i\geq 1}$ converges to 1. We conclude that the Gauss map is invariant with respect to the Minkowski question mark measure. We now prove ergodicity by showing that $([0,1),\mathcal{B},\mu_?,\mathcal{T})$ is isomorphic to a Bernoulli system.

Theorem 2.21. *The pair $(\mathcal{T},\mu_?)$ is ergodic.*

Proof. Let $(X,\mathcal{F}_b,\mu_b,T_b)$ be the Bernoulli system with $X=\mathbb{N}^{\mathbb{N}}$ and $\mathcal{F}_b$ the product σ-algebra on X. Furthermore, T_b is the Bernoulli shift and we define $p=(p_1,p_2,p_3,\cdots)$ as the infinite probability vector that induces the Bernoulli measure μ_b, where $p_n=2^{-n}$.

As continued fraction expansions are unique for irrationals, but not for rationals, we need to remove a suitable set of measure zero from X in order to define a proper isomorphism. As such, we define the set

$$B = \{a_1 a_2 a_3 \cdots \in X_b : \quad \exists N \text{ such that } a_N = 1 \text{ and } \forall n > N, a_n = 0\}.$$

and consequently define the isomorphism $\phi : [0,1) \setminus \{0\} \to X \setminus (B \cup \{000\cdots\})$ by

$$\phi([a_1, a_2, \cdots, a_n]) = a_1 a_2, \cdots, a_n.$$

Clearly, ϕ is a bijection. Next, we check properties of an isomorphism by showing that it holds on the cylinders. Let $A = \{x : x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n\} \in \mathcal{F}_b$. Then

$$\phi^{-1}(A) = \Delta(a_1, a_2, \cdots, a_n) \in \mathcal{B},$$

from which it follows that both ϕ and ϕ^{-1} are measurable. Furthermore, ϕ preserves the measures

$$\mu_?(\phi^{-1}(A)) = \mu_?(\Delta(a_1, a_2, \cdots, a_n)) = 2^{-(a_1+a_2+\cdots+a_n)} = 2^{-a_1}2^{-a_2}\cdots 2^{-a_n} = \mu_b(A).$$

Lastly, let $[a_1, a_2, a_3, \cdots]$ denote the (possibly finite) continued fraction expansion of an arbitrary $x \in (0,1)$. Then as both $\mathcal{T}$ and T_b are left shifts, we have

$$(\phi \circ \mathcal{T})(x) = \phi([a_2, a_3, \cdots]) = a_2 a_3 \cdots = T_b(a_1 a_2 a_3 \cdots) = (T_b \circ \phi)(x).$$

Therefore, ϕ preserves the dynamics and we conclude that $([0,1), \mathcal{B}, \mu_?, \mathcal{T})$ and $(X, \mathcal{F}_b, \mu_b, T_b)$ are isomorphic. As the latter is a Bernoulli system, we conclude dat $(\mathcal{T}, \mu_?)$ is ergodic. □

Corollary 2.22. The Minkowski question mark is singular with respect to the Lebesgue measure.

Proof. As both $(\mathcal{T}, \mu_?)$ and $(\mathcal{T}, \gamma)$ are ergodic, it follows from Theorem 2.10 that $\mu_? = \gamma$ or $\mu_? \perp \gamma$. Suppose that $\mu_? = \gamma$ and consider the fundamental interval $\Delta(1) = [\frac{1}{2}, 1)$. Then

$$\gamma(\Delta(1)) = \frac{1}{\log 2}\int_{1/2}^{1} \frac{1}{1+x}dx = \frac{\log \frac{4}{3}}{\log 2} \neq \frac{1}{2} = ?(1) - ?(1/2) = \mu_?(\Delta(1)),$$

which contradicts the assumption that $\mu_? = \gamma$. Hence we conclude that the Minkowski question mark and the Gauss measure are singular. As the Gauss measure and Lebesgue measure are equivalent, it follows from basic measure theory that Minkowski question mark and Lebesgue measure are singular as well. □

Recall that if two measures are singular, they have a different support. Intuitively, this means that singularity of measures tells us that we cannot compare these measures. However, by constructing isomorphisms, we can show that two measure preserving systems can exhibit the same dynamical behaviour even when the difference in measure structure lies in the singularity of the measures. This shows that although two measures are singular and not comparable, their behaviour is comparable within some systems. We illustrate this with the following example.

Example 2.23. On each interval $[\frac{1}{2^n}, \frac{1}{2^{n-1}})$, $n \geq 1$, define the measure preserving transformation $S : [0, 1) \to [0, 1)$ by

$$Sx = 2^n x - 1.$$

Also, let $d(x) = n$ on $[\frac{1}{2^n}, \frac{1}{2^{n-1}})$ and $d_n(x) = d(S^{n-1}x)$. In particular, it follows that $d_1 x = d(x)$ and $Sx = 2^{d(x)}x - 1$. By rewriting these equations, we find

$$\begin{aligned} x &= \frac{1}{2^{d_1(x)}} + \frac{Sx}{2^{d_1(x)}} \\ &= \frac{1}{2^{d_1(x)}} + \frac{1}{2^{d_1(x)}}\Big(\frac{1}{2^{d_2(x)}} + \frac{Sx}{2^{d_2(x)}}\Big) \\ &= ... \\ &= \frac{1}{2^{d_1(x)}} + \frac{1}{2^{d_1(x)+d_2(x)}} + \cdots + \frac{S^n x}{2^{d_1(x)+d_2(x)+\cdots+d_n(x)}} \\ &= \sum_{i=1}^{\infty} \frac{1}{2^{d_1(x)+d_2(x)+\cdots+d_i(x)}}. \end{aligned}$$

Note the similarity with $?(x)$. We briefly show that $([0, 1), \mathcal{T}, \mathcal{B}, \mu_?)$ and $([0, 1), \mathcal{B}, \lambda, S)$ are isomorphic. Defining this isomorphism is similar to the one in the proof of ergodicity of $(\mathcal{T}, \mu_?)$. Let $[a_1, a_2, a_3, \cdots]$ be the - possibly finite - continued fraction expansion of an arbitrary $x \in [0, 1)$. Then define

$$\phi(x) = \phi([a_1, a_2, a_3, \cdots]) = \sum_{i=1}^{\infty} \frac{1}{2^{a_1+a_2+\cdots+a_i}},$$

which is clearly a bijection. Furthermore, as the dyadic intervals generate the Borel σ-algebra, we check the properties of an isomorphism by considering such an interval. Hence, let A be an arbitrary dyadic interval

$$A = \Big[\sum_{i=1}^{n} \frac{1}{2^{a_1+a_2+\cdots+a_i}}, \sum_{i=1}^{n} \frac{1}{2^{a_1+a_2+\cdots+a_i}} + \frac{1}{2^{a_1+a_2+\cdots+a_n}}\Big).$$

Then any $x \in A$ is given by

$$x = \frac{1}{2^{a_1}} + \frac{1}{2^{a_1+a_2}} + \cdots + \frac{1}{2^{a_1+a_2+\cdots+a_n}} + \cdots .$$

Moreover, we have that $\phi^{-1}(A) = \Delta(a_1, a_2, \cdots, a_n) \in \mathcal{B}$ and

$$\mu_?(\phi^{-1}(A)) = \mu_?(\Delta(a_1, a_2, \cdots, a_n)) = 2^{-(a_1+a_2+\cdots+a_n)} = \lambda(A).$$

Lastly,

$$(\phi \circ \mathcal{T})(x) = \phi([a_2, a_3, \cdots]) = \sum_{i=2}^{\infty} \frac{1}{2^{a_2+a_3+\cdots+a_i}} = 2^{a_1} \sum_{i=1}^{\infty} \frac{1}{2^{a_1+a_2+\cdots+a_i}} - 1$$

$$= S\Big(\sum_{i=1}^{\infty} \frac{1}{2^{a_1+a_2+\cdots+a_i}}\Big) = (S \circ \phi)(x).$$

We conclude that $([0,1), \mathcal{T}, \mathcal{B}, \mu_?)$ and $([0,1), \mathcal{B}, \lambda, S)$ are isomorphic.

The previous example shows an important property of the Minkowski question mark. That is, the Minkowski question mark linearises the Gauss map. As linear systems are simpler mathematical objects, this is a useful property. However, we do not need it in this thesis.

Summarising, the Minkowski question mark possesses some remarkable properties. Among other things, it is strictly monotone, continuous and singular. When we consider the system $(X, \mathcal{B}, \mu_?, \mathcal{T})$, we find that is both isomorphic to a Bernoulli and a linear system. This implies that the Gauss map is ergodic under the Minkowski question mark and the question mark linearises the Gauss map. That is, the diagram

$$\begin{tikzcd}
([0,1), \mu_?) \arrow[r, "\mathcal{T}"] \arrow[d, "\phi"] & ([0,1), \mu_?) \arrow[d, "\phi"] \\
([0,1), \lambda) \arrow[r, "S"] & ([0,1), \lambda)
\end{tikzcd}$$

commutes. We use the Minkowski question mark to introduce a notion of normality for the continued fraction expansion in chapter 4, which refer to as Minkowski normality. More importantly, in Theorem 4.13, we prove Minkowski normality of a constructed number $\mathcal{K}$, see (4.10). First, however, we discuss normality results that have been obtained thus far.

Chapter 3

Normal numbers and related developments

Numbers can be represented in numerous ways. Examples of representations include binary representations, decimal representations and continued fraction expansions. Each such representation corresponds to a finite or infinite sequence of digits that make up such representation. When regarding these digits as random variables, we can consider the distribution of the sequence of digits - if it exists. Normality of a number is then characterised as a distribution property of the (infinite) sequence of digits that corresponds to this number. In this section, we discuss several types normality and results that have been obtained since Borel introduced the notion of normal numbers in 1909. The results contain constructions of normal numbers as well as proofs of existence. The goal of discussing these constructions and existence results, is to provide a historical framework on results on normal numbers, where the importance lies with the techniques that have been used to obtain the construction results. Roughly speaking, there are two techniques that are used when proving normality of a constructed number. We introduce these techniques by considering two results in more detail. The first technique is shown by going over a result from David Champernowne, who was the first to explicitly construct a normal number. The second technique is due to Abram Besicovitch, which was used by Copeland and Erdös to prove the first generalised construction. Both of these results are discussed in the next section, where we focus on a type of normality as introduced by Borel. The section thereafter discusses a type of normality that is related to continued fraction expansions, after which we discuss other normality results and provide some final remarks.

3.1.0 *Constructions of numbers normal in a base*

In 1909 Émile Borel introduced the notion of normality [16]. The type of normality that he introduced is related to a generalised form of decimal expansions. In this section, we define this type of normality and discuss important results that have been obtained so far. Among these results are those of Champernowne, Copeland and Erdös, which are treated in more detail. The results that are discussed in this section are all related to the same type of normality, which we define as follows.

Definition 3.1 (Normal in a base)**.** Given an integer $b \geq 2$, an irrational number $x = \sum_{i=1}^{\infty} \frac{a_i}{b^i} = 0.a_1a_2a_3\cdots \in [0,1)$ is called normal in base b if for any $k \geq 1$ and any block $d = d_1d_2\cdots d_k$ with $d_i \in \{0,1,\cdots,b-1\}$, one has

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{a_ia_{i+1}\cdots a_{i+k-1}=d_1d_2\cdots d_k\}} = \frac{1}{b^k}. \tag{3.1}$$

If the above holds for $k=1$, then we say that x is *simply normal* in base b. Furthermore, when x is normal in all bases $b \in \mathbb{N}$, we say that x is *absolutely normal.*

In the above type of normality, x is written in its so-called *b-adic expansion*. The digits of the b-adic expansion are generated by iterations of the map $Rx = bx \mod 1$, which is a left shift for the b-adic expansion. Analogous to the proof of Theorem 2.21, one can prove that the sytem $(X, \mathcal{B}, \lambda, R)$ is ergodic. This and other results on this type of expansion can for instance be found in [11, Example 3.1.2]. Furthermore, (3.1) implies that the asymptotic frequency of occurrences of d in x is equal to b^{-k}. Lastly, note that $b = 2$ and $b = 10$ correspond to the binary, respectively, decimal expansion of a number x.

Remarkably, it turns out that almost all numbers are absolutely normal. This was almost proved by Émile Borel in 1909. He almost proved the theorem in the sense that he assumed one of the in-between steps rather than proving it, leaving a gap in the proof. A year later Faber [17] filled the gap and therefore the proof was concluded. The key observation in Borel's proof is that the digits of the b-adic expansion, seen as random variables, are i.i.d. uniformly distributed. An argument similar to the strong law of large numbers then proves that almost all numbers are simply normal. Subsequently, a similar analysis can be applied to blocks of length k, proving normality. Using ergodic theory, one can also give a more direct proof.

Theorem 3.2. *λ almost every number in $[0, 1)$ is absolutely normal.*

Proof. Let $x \in [0, 1)$ and consider the ergodic system $(X, \mathcal{B}, \lambda, R)$. Then for any $k \geq 1$ and any block $d = d_1 d_2 \cdots d_k$, $d_i \in \{0, 1, \cdots, b-1\}$, it follows from Theorem 2.6 that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{C_d}(R^i x) = \lambda(C_d) = \frac{1}{b^k} \qquad a.e.,$$

where $C_d = [\frac{d_1}{b} + \frac{d_2}{b^2} + \cdots + \frac{d_k}{b^k}, \frac{d_1}{b} + \frac{d_2}{b^2} + \cdots + \frac{d_k+1}{b^k})$ is the cylinder set containing d. Moreover, as a countable union of sets of measure zero has measure zero, it follows that almost all numbers are absolutely normal. □

In order to see that the asymptotic average in the proof is equivalent with the asymptotic condition for normality, we refer to the last paragraph of section 2.1. The cylinder set C_d is the set of numbers of which the first k digits are specified by d. Hence, as R is a left shift and $\mathbb{1}_{C_d}(R^i x) = 1$ if and only if $R^i x \in C_d$, we count the number of times that the first k digits of $R^i x$ coincide with d, $i \geq 1$. Therefore, the asymptotic average in the proof is the asymptotic frequency of occurrences of d in x, which is precisely the left hand side in (3.1).

In order to prove normality of an irrational number, one needs to know the corresponding infinite expansion or recognise a pattern in its expansion. Apart from the fact that in most cases it is practically impossible to consider an infinite sequence of digits, irrational numbers usually do not exhibit repetition and if patterns exist, they are hard to detect. Hence, for the time being, a construction is the only way to explicitly exhibit a normal number. Many constructions have been developed, both of normal numbers as introduced by Borel, as for other types of normality that have been introduced after 1909. The most well-known construction of a normal number is due to David Champernowne. He proved that the decimal number that is obtained by concatenating the natural numbers is normal in base 10. That is, Champernowne proved that the number

$$\mathcal{C}_{10} := 0.\,1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12\,13\,14\,\cdots, \tag{3.2}$$

is normal in base 10. In order to prove this, he first proves two more general theorems. In what follows, we state and prove these theorems. Moreover, we follow [9] and provide details of the proofs. The key takeaway is the technique that is used in the proofs, which is based on asymptotics and combinatorics.

3.1.1 Champernowne and counting for normality

Let $0.S$ be a decimal and denote its sequence of digits by S. Then $0.S$ is normal in base 10 if for any $k \geq 1$ and any block $d = d_1 d_2 \cdots d_k$ with $d_i \in \{0, ..., 9\}$, the asymptotic relative frequency of d in S is 10^{-k}. Define $G_n(S, d)$ as the number of occurrences of d in the first n digits of S. Then, $0.S$ is normal in base 10 if and only if

$$G_n(S, d) = 10^{-k} n + o(n), \tag{3.3}$$

as $n \to \infty$. We next introduce some other definitions and notation

- $|y|$ denotes the number of digits in the sequence y;
- s_r denotes the sequence of 10^r possible permutations of r digits, concatenated in lexicographical ordering. For reasons of convenience, we place comma's between the different elements of s_r. E.g. s_2 = 00,01,02,…, 19, 20, 21, …, 92, 93, 94, 95, 96, 97, 98, 99. Also, note that there are $10^r - 1$ commas within s_r and that $|s_r| = r10^r$;
- S_r and S denote the finite, respectively infinite, sequences, $s_1 s_2 \cdots s_r$ and $s_1 s_2 s_3 \cdots$. Note that $|S_r| = \sum_{i=1}^{r} |s_i| = \sum_{i=1}^{r} i10^i$.

Using the above, we prove the following theorems:

Theorem 3.3. *[9, Theorem I] Let s_r be defined as above. Then $0.S = \cdot s_1 s_2 s_3 \cdots$ is normal in base 10.*

Theorem 3.4. *[9, Theorem II] Let s_r be defined as above and, for $\rho \in \mathbb{N}$, let ${}_\rho s_r$ denote the sequence that is formed by concatenating s_r ρ times. Then $0._\rho S = 0._\rho s_{1\rho} s_{2\rho} s_3 \cdots$ is normal in base 10.*

Theorem 3.5. *[9, Theorem III] C_{10} is normal in base 10.*

Note that Theorem 3.3 is a specific case of Theorem 3.4, namely with $\rho = 1$. We will show that normality of Champernowne's number follows almost directly from Theorem 3.4. The proofs of these theorems are quite elegant and rely mostly on combinatorial arguments. The idea is to count the number of occurrences of an arbitrary block d (of finite length) in s_r and extend this to the number of occurrences in S_r. Subsequently, we count the number of occurrences of d in the first n digits of s_r, which is then used to count the number of occurrences in S. Another key aspect of the proof is to distinguish the following cases.

- The block d occurs without any comma between its digits. In this case we say that d occurs *undivided.*
- The block d occurs with a comma between two of its digits. In this case we say that d occurs *divided.*

For example, suppose $d = 92$ and we consider the occurrences of d in s_2. Then d occurs undivided in the element 92 and divided at $\ldots, 19, 20, 21, \ldots$. We are now ready to give proof of Theorem 3.3.

Proof of Theorem 3.3. Let d be a block of length k and consider an arbitrary permutation of length r, i.e. an arbitrary element of s_r. For $r < k$, the block d cannot occur undivided. However, if $r \geq k$, the first digit of d can occur on any of the first $r - k + 1$ positions. Subsequently, the remaining $r - k$ digits can be chosen in 10^{r-k} ways. Hence d can occur exactly in $(r - k + 1)10^{r-k}$ ways.

Next, we give an upper bound for the number of ways that d can occur divided in the sequence s_r. As s_r consists of 10^r elements, there are $10^r - 1$ commas within s_r. Similarly, as d consists of k digits, there are $k - 1$ positions at which d can be divided. Thus, there are exactly $(k - 1)(10^r - 1)$ ways that d can occur divided.

From the above we conclude that

$$\begin{aligned} G_{|s_r|}(s_r, d) &= (r - k + 1)10^{r-k} + \mathcal{O}(10^r) \\ &= 10^{-k} r 10^r + (-k + 1)10^{-k}10^r + \mathcal{O}(10^r) \\ &= 10^{-k}|s_r| + o(|s_r|), \end{aligned}$$

as $r \to \infty$. We extend this to an estimate for $G_{|S_r|}(S_r, d)$.

$$\begin{aligned} G_{|S_r|}(S_r, d) &= \sum_{i=1}^{r} G_{|s_i|}(s_i, d) + \mathcal{O}(r) \\ &= \sum_{i=1}^{r} 10^{-k}|s_i| + o(|s_i|) + \mathcal{O}(r) \\ &= 10^{-k} \sum_{i=1}^{r} |s_i| + o(|s_r|) \\ &= 10^{-k}|S_r| + o(|S_r|), \end{aligned}$$

as $r \to \infty$. Note that the term $\mathcal{O}(r)$ results from the possibility that d can occur divided between consecutive elements of S_r, namely s_i and s_{i+1} for $i = 1, ...r - 1$. The next step in the proof is to count the number of occurrences of a block d within the first n digits of s_r, which we then extend to the number of occurrences in the first n digits of S. We consider the number of undivided occurrences of d in the first n digits of s_r. As s_r is the concatenation of all possible permutations of r digits (in lexicographical order), we may suppose that the n-th digit of s_r occurs within an element $p_{r-1}p_{r-2}\cdots p_0$ of s_r, where $p_i \in \{0,\dots,9\}$ for $i = 0,\dots,r-1$. As the elements of s_r are ordered lexicographically, we can express the n-th digit as

$$n = r\sum_{i=0}^{r-1} p_i 10^i + \theta r, \qquad 0 < \theta \leq 1. \tag{3.4}$$

The intuition behind this is that, due to the lexicographical ordering, for every p_i the subsequent i digits (within the element of s_r) $p_{i-1}\cdots p_0$ can be chosen in 10^i ways. Summing over all possible j's, we get a count of the number of elements of s_r that preceed the one with the n-th digit. As each element of s_r is a sequence of r digits, (3.4) correctly expresses n. Now let $G_{n,j}(s_r, d)$ denote the number of times that d can occur undivided in the first n digits of s_r such that the first digit of d is the j-th digit of an element in $s_r = p_{r-1}\cdots p_0$. I.e. $d_1 = p_{r-j-1}$. Trivially, if $j > r-k+1$, then $G_{n,j}(s_r,d) = 0$. If $j \leq r-k+1$ and we fix the position of d in an element of s_r, then we can choose the successive $r-k-j+1$ digits of the element in $10^{r-k-j+1}$ ways. The $j-1$ digits preceding d can be chosen in either

$$\sum_{i=r-j+1}^{r-1} p_i 10^{i+j-r-1} \qquad \text{or}$$
$$\sum_{i=r-j+1}^{r-1} p_i 10^{i+j-r-1} + 1$$

ways. The argument here is similar to the one used for (3.4), the difference being that we cannot freely choose all subsequent i elements $p_{i-1}\cdots p_0$, but only the digits $p_{i-1}\cdots p_{r-j}$ for $i = r-j+1,\dots r-1$. This amounts to a total of $(i-1)-(r-j) = i+j-r-1$ digits. We conclude that the number of undivided occurrences of d in the first n digits of s_r is given by

$$\begin{aligned} G_{n,j}(s_r,d) &= 10^{r-k-j+1}\Big(\sum_{i=r-j+1}^{r-1} p_i 10^{i+j-r-1} + \theta'\Big) \\ &= 10^{-k}\Big(\sum_{i=r-j+1}^{r-1} p_i 10^{i} + \theta' 10^{r-j+1}\Big) \qquad\qquad 0 \leq \theta' \leq 1. \end{aligned}$$

By summing over all possible j's and taking into account the number of divided occurrences, we can extend this to $G_n(s_r,d)$. The number of divided occurrences is $\mathcal{O}(10^r)$. Hence, as $r\to\infty$, we find that

$$\begin{aligned} G_n(s_r,d) = \sum_{j=1}^{r-k+1} G_{n,j}(s_r,d) &= 10^{-k}\sum_{j=1}^{r-k+1}\Big(\sum_{i=r-j+1}^{r-1} p_i 10^i\Big) + \mathcal{O}(10^r) \\ &= 10^{-k}\sum_{i=k}^{r-1}(i+1-k)p_i 10^i + \mathcal{O}(10^r) = 10^{-k} r\sum_{i=0}^{r-1} p_i 10^i + \mathcal{O}(10^r) \end{aligned}$$

$$= 10^{-k}n + o(|s_r|). \tag{3.5}$$

In order to prove normality of S, we suppose that the n-th digit of S occurs as the m-th digit of s_r. That is, $n = |S_{r-1}| + m = \sum_{i=1}^{r-1}|s_i| + m$. It follows from (3.5) that, as $n \to \infty$, the number of occurrences of d in the first n digits of S is given by

$$\begin{aligned}
G_n(S,d) &= G_{|S_{r-1}|}(S_{r-1},d) + G_y(s_r,d) + \mathcal{O}(1)\\
&= \sum_{i=1}^{r-1} G_{|s_i|}(s_i,d) + G_m(s_r,d) + \mathcal{O}(1)\\
&= \sum_{i=1}^{r-1} 10^{-k}|s_i| + 10^{-k}m + o(|s_r|)\\
&= 10^{-k}\Big(\sum_{i=1}^{r-1}|s_i| + m\Big) + o(|s_r|)\\
&= 10^{-k}n + o(n).
\end{aligned}$$

We conclude that $0.S$ is normal in base 10. □

Champernowne's original work provides a short proof of the normality of $0._\rho S$, as it follows almost directly from Theorem 3.3. In the proof below we follow his work and provide details.

Proof of Theorem 3.4. Recall that ${}_\rho s_r$ is defined as the sequence that is obtained by concatenating s_r ρ times. Now define ${}_\rho S_r$ as the sequence ${}_\rho s_1 {}_\rho s_2 \cdots {}_\rho s_r$ and notice that

- $|{}_\rho s_r| = \rho|s_r|$;
- $|{}_\rho S_r| = \rho|S_r| = \rho\sum_{i=1}^{r}|s_i|$.

By differentiating between occurrences of d within s_r and d occurring divided over two consecutive repetitions of s_r, we find that the number of occurrences of d in ${}_\rho s_r$ is ρ times the number of occurrences of d in s_r plus some term that disappears in the little-o term.

$$G_{|{}_\rho s_r|}({}_\rho s_r, d) = \rho G_{|s_r|}(s_r,d) + \rho(k-1) + o(|s_r|) = \rho 10^{-k}|s_r| + o(|s_r|). \tag{3.6}$$

Similar to the last step of the previous proof, suppose that the n-th digit of ${}_\rho S$ is the M-th digit of ${}_\rho s_r$ and subsequently suppose that this is the m-th digit of one of the repetitions of s_r. Then for some $0 \le \alpha < \rho$ we have that

$$n = |{}_\rho S_{r-1}| + M = |{}_\rho S_{r-1}| + \alpha|{}_\rho s_r| + m = \rho\sum_{i=1}^{r-1}|s_i| + \alpha\rho|s_r| + m. \tag{3.7}$$

Then, using (3.5), (3.6) and (3.7) we find that, as $n \to \infty$, the number of occurrences of d in the first n digits of ${}_\rho S$ is given by

$$\begin{aligned}
G_n({}_\rho S, d) &= G_{|{}_\rho S_{r-1}|}({}_\rho S_{r-1}, d) + \alpha G_{|{}_\rho s_r|}({}_\rho s_r, d) + G_m(s_r, d) \\
&= \sum_{i=1}^{r-1} G_{|{}_\rho s_i|}({}_\rho s_i, d) + \alpha G_{|{}_\rho s_r|}({}_\rho s_r, d) + G_m(s_r, d) \\
&= \sum_{i=1}^{r-1} \rho 10^{-k}|s_i| + \alpha\rho 10^{-k}|s_r| + 10^{-k}m + o(|s_r|) \\
&= 10^{-k}\Big(\rho \sum_{i=1}^{r-1} |s_i| + \alpha\rho|s_r| + m\Big) + o(|s_r|) \\
&= 10^{-k}n + o(n),
\end{aligned}$$

Hence $0.{}_\rho S$ is normal in base 10. $\square$

In order to prove normality of what is now known as Champernowne's constant, Champernowne first proves that the decimal $0.{}_9 S' = \cdot 101112\cdots$ is normal in base 10, from which it follows that C_{10} is normal in base 10 as well.

Proof of Theorem 3.5. The number $0.{}_9 S'$ is obtained by taking the sequence ${}_9 S$ and inserting 1 digit "after each comma". As $0.{}_9 S$ is normal in base 10 by Theorem 3.4 we can prove normality of $0.{}_9 S'$ by showing that the insertion of the extra digits does not influence this property.

Let $C(n)$ denote the number of commas in the first n digits of ${}_9 S$ and suppose that the n-th digit of ${}_9 S$ occurs within some s_r. Then it follows that $C(n) = \mathcal{O}(10^r) = o(n)$, as $n \to \infty$. By inserting the extra digits, the n-th digit of ${}_9 S$ becomes the n'-th digit of ${}_9 S'$. Hence

$$n' = n + C(n) + \mathcal{O}(1) = n + o(n),$$

where the $\mathcal{O}(1)$ term results from the case that the n'-th digit is one of the insertions. Also note that inserting a digit after the commas can only alter the number of divided occurrences. Thus the number of occurrences within the first n digits of ${}_9 S$ is altered by a maximum of $kC(n)$. Therefore, we find that

$$G_{n'}({}_9 S', d) = G_n({}_9 S, d) + \mathcal{O}(C(n)) = 10^{-k}n + o(n) = 10^{-k}n' + o(n').$$

We conclude that $0.{}_9 S'$ is normal in base 10. Hence so is C_{10}. $\square$

Without giving a proof, Champernowne states theorems about the normality of various numbers and conjectures that the number obtained by the sequence of primes is also normal in base 10. He argues that normality of these numbers can be proved through techniques similar to those used in his paper, which is based on a combination of combinatorics and asymptotics.

Throughout the years, others have been able to prove theorems that provide us with generalised constructions of normal numbers. The first to do this were Copeland and Erdös [10]. In the proof they use a concept that is widely used when proving normality of a number. Therefore, we go over their work and provide details of the proof.

3.1.2 Copeland, Erdös and Besicovitch's (ϵ, k)-normality

Copeland and Erdös [10] provided the first generalised construction of a normal number. A key concept that is used in the proofs is that of (ϵ, k)*-normality*. This concept was introduced by Besicovitch in 1935 [5], who used it to prove normality of the decimal that is formed by concatenating the squares of the natural numbers. Since then, the concept has frequently been used for constructing or proving the existence of normal numbers. It is defined as follows.

Definition 3.6 ((ϵ, k)-normality)**.** [10, Definition] Given a base $b \geq 2$ and $\epsilon > 0$, a number $A = a_1 a_2 \cdots a_n$ is said to be (ϵ, k)-normal in base b, if for any $k \geq 1$ and any block $d = d_1 d_2 \cdots d_k$ with $d_i \in \{0, 1, \ldots, b-1\}$, the relative frequency of d in A is between $b^{-k} - \epsilon$ and $b^{-k} + \epsilon$. That is,

$$b^{-k} - \epsilon \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{a_i a_{i+1} \cdots a_{i+k-1} = d_1 d_2 \cdots d_k\}} \leq b^{-k} + \epsilon. \tag{3.8}$$

In other words, a number is (ϵ, k)-normal if it is approximately normal. Using this concept, Copeland and Erdös prove the following theorem.

Theorem 3.7. *[10, Theorem] If $(a_i)_{i \geq 1}$ is an increasing sequence of integers such that for every $\theta < 1$ and sufficiently large N, we have that $\sum_{i \geq 1} \mathbb{1}_{\{a_i \leq N\}} > N^{\theta}$, then*

$$0.a_1 a_2 a_3 \cdots$$

is normal in the base b in which the integers a_i are expressed, $i \geq 1$.

The proof of Theorem 3.7 is given by showing that the number of a_i's in the sequence $(a_i)_{i \geq 1}$ that are smaller or equal than N and are not (ϵ, k)-normal, is of order o(1) as $N \to \infty$. In other words, they prove that almost all a_i's are (ϵ, k)-normal. First, however, they prove that the number of a_i's that are not (ϵ, k)-normal is bounded. This is captured in the lemma below.

Lemma 3.8. *[10, Lemma] The number of integers up to N, for sufficiently large N, which are not (ϵ, k)-normal in a given integer base b is less than N^{δ}, where $\delta = \delta(\epsilon, k, b) < 1$.*

We now give an outline of Theorem 3.7 and Lemma 3.8. Here we follow Copeland and Erdös [10] and provide details of their proof.

Proof of Theorem 3.7 and Lemma 3.8. The lemma is first proved for $(\epsilon, 1)$-normality, which is then extended to (ϵ, k)-normality. Copeland and Erdös note that the number of a_i's up to N that do not have the right frequency is at most

$$b \sum_{K < \frac{(1-\epsilon)n}{b}} (b-1)^{n-K} \binom{n}{K} + b \sum_{K > \frac{(1+\epsilon)n}{b}} (b-1)^{n-K} \binom{n}{K}, \tag{3.9}$$

where n is such that $b^{n-1} \leq N < b^n$. In other words, n is such that $|N| \leq n$. To see that the above holds true, we first recall that $(\epsilon, 1)$-normality means that the relative frequency of a single digit should be between $b^{-1} - \epsilon$ and $b^{-1} + \epsilon$. This is equivalent to saying that K, the number of occurrences of a single digit, should be between $\frac{(1-\epsilon)n}{b}$ and $\frac{(1+\epsilon)n}{b}$. Therefore, when a number is not $(\epsilon, 1)$-normal, there are two cases.

- The number of occurrences is lower than $\frac{(1-\epsilon)n}{b}$. This corresponds to the values over which we sum in the first term in (3.9).
- The number of occurrences is higher than $\frac{(1+\epsilon)n}{b}$. This corresponds to values over which we sum in the second term in (3.9).

Hence K denotes the number of occurrences of a digit $\alpha \in \{0, 1, \ldots, b-1\}$ in a number that has at most n digits. Therefore, there are at most $n - K$ other digits that have to be different from α. There are $b - 1$ choices for the latter, which results in $(b-1)^{n-K}\binom{n}{K}$ possibilities. As there are b choices for α, we find that (3.9) correctly bounds the total number of a_i's up to N that do not have the right frequency. Due to the fact that the binomial coefficient is increasing and then decreasing in K, it attains a maximum. This maximum is used to bound (3.9), which then proves the lemma for $(\epsilon, 1)$-normality. The argument extending this to (ϵ, k)-normality is more or less identical. The difference is that the digits of a number $m \leq N$ are grouped in groups of size k, which can be interpreted as a single digit expressed in base b^k. With this, we conclude the outline for the proof of Lemma 3.8. We now describe the last few steps that are made in order to prove the main theorem.

We know that there are at least N^θ numbers in the sequence $(a_i)_{i\geq 1}$ that are smaller or equal to N. Of these numbers, there are at most $b^{n(1-\epsilon)} \geq N^{1-\epsilon}$ numbers that have $n(1-\epsilon)$ digits. Therefore, at least $N^\theta - N^{1-\epsilon}$ of the numbers up to N have at least $n(1-\epsilon)$ digits. This amounts to a total of $n(1-\epsilon)(N^\theta - N^{1-\epsilon})$ digits. Let $m = \sup\{i : a_i \leq N\}$. Then it follows from Lemma 3.8 that the number of a_i's up to a_m that are not (ϵ, k)-normal is bounded by N^δ. Hence we find that

$$\begin{aligned}
b^{-k} - \epsilon + \frac{(n-k+1)N^\delta}{n(1-\epsilon)(N^\theta - N^{1-\epsilon})} &< \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\{a_i a_{i+1}\cdots a_{i+k-1} = d_1 d_2 \cdots d_k\}} \\
&< b^{-k} + \epsilon + \frac{(n-k+1)N^\delta}{n(1-\epsilon)(N^\theta - N^{1-\epsilon})}; \\
b^{-k} - \epsilon + \frac{N^{\delta-\theta}}{(1-\epsilon)(1 - N^{1-\epsilon-\theta})} &< \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\{a_i a_{i+1}\cdots a_{i+k-1} = d_1 d_2 \cdots d_k\}} \\
&< b^{-k} + \epsilon + \frac{N^{\delta-\theta}}{(1-\epsilon)(1 - N^{1-\epsilon-\theta})}.
\end{aligned}$$

Now, taking θ greater than δ and greater than $1 - \epsilon$, it follows that

$$b^{-k} - \epsilon < \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{a_i a_{i+1}\cdots a_{i+k-1} = d_1 d_2 \cdots d_k\}} < b^{-k} + \epsilon.$$

Hence we conclude that $0.a_1a_2a_3\cdots$ in base b. $\square$

The proof of Champernowne's conjecture that the decimal obtained by the sequence of primes is normal in base 10 follows from the fact that for any $c < 1$ and sufficiently large N, the number of primes up to N is bounded by $\frac{cN}{\log N}$. On a more general note, Copeland and Erdös show that almost all numbers are (ϵ, k)-normal. Moreover, their result implies that we expect to get a normal number whenever we take a sufficiently dense subset of the positive integers and concatenate them in increasing order [34]. Next, we recap the two techniques and discuss further generalisations for normality in a base.

3.1.3 Further generalisations of normality in a base

Since Borel introduced the notion of normality, many results have been obtained for normality in a base. Most results have been obtained through a proof that either relies on counting or on (ϵ, k)-normality. Counting proofs generally try to give a direct proof that the asymptotic frequency of occurrences of an arbitrary block is indeed the frequency that is associated with normality. However, proofs that use (ϵ, k)-normality generally try to show that the parts of the sequence of digits that do not have the right frequency are negligible. By looking at parts the parts that are not (ϵ, k)-normal, we are looking at some form of discrepancy. Given a base $b \in N_{\geq 2}$, the discrepancy in the first n digits of $x = \sum_{i=1}^{\infty} \frac{a_i}{b^i}$ is defined as

$$D_n(x) = \sup_{d_1d_2\cdots d_k \in \{0,1,\ldots,b-1\}^k} \Big\{ \Big| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{a_i a_{i+1}\cdots a_{i+k-1} = d_1 d_2 \cdots d_k\}} - \frac{1}{b^k} \Big| \Big\}.$$

Thus the discrepancy can be interpreted as a number's (maximum) deviation from normality. Note that it follows that x is normal if and only if $D_n(x) \to 0$. This technique is also frequently used in literature. However, as it is an analogue to (ϵ, k)-normality, we do not treat it in further detail. We next discuss further generalisations, after which we introduce the type of normality that is associated with continued fractions.

Besides proving the first generalised construction of a normal number, Copeland and Erdös conjectured that for any polynomial $f(x)$, the number $0.f(1)f(2)f(3)\cdots$ would be normal in base 10. This conjecture was partially proved by Davenport and Erdös.

Theorem 3.9. *[12, Theorem 1] Let $f(x)$ be any non-constant polynomial in x that attains integer values for $x \in \mathbb{N}$. Then $0.f(1)f(2)f(3)\cdots$ is normal in base 10.*

This theorem is another step in generalising the results mentioned so far. For instance, normality of Champernowne's number follows from $f(x) = x$ and the result from Besicovitch from $f(x) = x^2$. This result was generalised even further by Nakai and Shiokawa, who used discrepancy estimates to prove that the above theorem also holds when we allow the polynomial to attain non-integer values. Moreover, they proved the following theorem, which generalises all construction results mentioned so far.

Theorem 3.10. *[26, Corollary] Let $f(x)$ be any real-valued, non-constant polynomial such that $f(x) > 0$ for $x > 0$. Then*

$$0.\lfloor f(1)\rfloor \lfloor f(2)\rfloor \lfloor f(3)\rfloor \cdots,$$

with $\lfloor f(n)\rfloor$ expressed in base $b \in \mathbb{N}$ for all n, is normal in the base b.

Summarising, there are quite some results on numbers that are normal in a base. The most well-known concrete example of a normal number is $\mathcal{C}_{10}$, the base 10 Champernowne number. Normality of $\mathcal{C}_{10}$ is initially proved by a combination of combinatorics and asymptotics, but also follows from generalised construction results. Furthermore, we introduced the concept of (ϵ, k)-normality, which is used to prove generalised constructions of numbers normal in a base. The importance of this concept is also apparent in the next section, where we consider a type of normality that is associated to continued fraction expansions.

3.2.0 *Constructions of continued fraction normal numbers*

Thus far we have considered b-adic expansions and the distribution of the corresponding sequence $(a_i)_{i\geq 1}$ under the Lebesgue measure. Next, we discuss normality results related to continued fraction expansions. The digits of this expansion can be studied using the Gauss map, which is invariant with respect to the Gauss measure. Moreover, let Δ_k be an arbitrary fundamental interval of order k. Then Gauss proved that the Lebesgue measure of the set $\mathcal{T}^{-n}\Delta_k$ converges weakly to the Gauss measure of Δ_k, as $n \to \infty$. The type of normality that is naturally associated to continued fraction expansions is therefore defined as follows.

Definition 3.11 (Continued fraction normality)**.** We say that $x \in [0, 1)$ is *continued fraction normal*, if for any $k \geq 1$ and any block $d = d_1, d_2, \cdots, d_k$, $d_i \in \mathbb{N}$ we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{a_i, a_{i+1}, \cdots, a_{i+k-1} = d_1, d_2, \cdots, d_k\}} = \gamma(\Delta(d)). \tag{3.10}$$

Alternatively, we can rewrite (3.10) in terms of the Gauss map and fundamental intervals. That is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\Delta(d)}(\mathcal{T}^i x) = \gamma(\Delta(d)).$$

Analogous to the proof that almost all numbers are normal, one can also prove that almost all numbers are continued fraction normal. As the Lebesgue and Gauss measure are equivalent, see (2.17), this implies that almost all numbers are continued fraction normal Lebesgue almost everywhere.

Contrary to the number of results for normality in a base, the number of results for this type of normality are limited. So far, there are three construction results. The first construction was due to Postnikov [29], who used Markov chains to construct a continued fraction normal number. The construction is such that each element of the Markov chain, is a long finite block of digits that has approximately the right frequency [35]. Moreover, every next block of the sequence better approximates the desired frequency. These infinite number of blocks are then concatenated to form a number that is continued fraction normal. Another, more explicit, construction of a continued fraction normal number is due to Adler, Keane and Smorodinsky [1]. In 1981, Adler, Keane and Smorodinsky constructed a continued fraction normal number by concatenating continued fraction expansions of rationals in $[0, 1)$ [1]. They concatenated the continued fraction expansions of the following sequence of rationals

$$\frac{1}{2}, \quad \frac{1}{3}, \frac{2}{3}, \quad \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \quad \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \quad \ldots, \frac{n-1}{n}, \quad \frac{1}{n+1}, \cdots. \tag{3.11}$$

In other words, for each $n \in \mathbb{N}$, they constructed a sequence of rationals by taking all non-reduced fractions with denominator n in increasing order and writing these down increasing in n. Subsequently, they constructed a continued fraction normal number by concatenating the continued fraction expansions of the resulting sequence of rationals (3.11). This number is given by

$$x_{aks} = [2, \quad 3, 1, 2, \quad 4, 2, 1, 3, \quad 5, 2, 2, 1, 1, 2, 1, 4, \cdots] \approx 0.44034.$$

Theorem 3.12. *[1, Theorem] The number x_{aks} is continued fraction normal.*

The proof is similar to that of Copeland and Erdös [10]. The initial definition of (ϵ, k)-normality, however, is not suited for continued fractions. One reason for this, is that the continued fraction expansion is a type of number expansions that is not related to any base. Furthermore, the type of expansion also determines the type of distribution that is naturally associated to the expansion. In the case of the continued fraction expansion, this distribution is the Gauss measure, see section 2.3. As such, Adler, Keane and Smorodinsky introduce a continued fraction analogue of (ϵ, k)-normality called *m-good*. A rational is called m-good if its partial quotients are approximately distributed according to the Gauss measure. They show that nearly all rationals with denominator at most m are m-good and that they can neglect those that are not. Thus by concatenating the corresponding continued fraction expansions of all these rationals, they obtain an infinite continued fraction that has the desired frequency.

For a long time, the papers of Postnikov and Adler, Keane and Smorodinsky have been the only ones that contain constructions of continued fraction normal numbers. For each work, it took about 30 years before it was generalised. The generalisation of Postnikov's construction is due to Madritsch and Mance [24], which we discuss in the next section. Both of these works do not include a concrete constructed number that is continued fraction normal. This is different from the work of Adler, Keane and Smorodinsky and the generalisation of their work, which is due to Joseph Vandehey. Among other things, Vandehey proves that some explicit subsequences of (3.11) can be used to construct a continued fraction normal number. Moreover, he proves the following general theorem, which he also uses to give concrete constructions.

Theorem 3.13. *[35, Theorem 1.1] Let $(r_i)_{i\geq 1}$ denote the sequence of reduced fractions in the* *interval* $(0,1)$ *ordered in the following way*

$$r_1 = \frac{1}{2}, \quad r_2 = \frac{1}{3}, \quad r_3 = \frac{1}{3}, \quad r_4 = \frac{1}{4}, \quad r_5 = \frac{3}{5}, \quad \dots .$$

Let $f : \mathbb{N} \to \mathbb{N}$*, and define the number* x_f *as the number constructed by concatenating the continued fraction expansions of the rationals* $r_{f(1)}, r_{f(2)}, r_{f(3)}, \dots$*. Let* $L(r)$ *denote the length of the continued fraction expansion of r. Suppose that*

$$N = o\Big(\sum_{i=1}^{N} L(r_{f(i)})\Big) \quad \textit{and} \quad \max_{1\le i\le N} L(r_{f(i)}) = \mathcal{O}\Big(\frac{1}{N}\sum_{i=1}^{N} L(r_{f(i)})\Big)$$

and that for any $S \subset \mathbb{N}$ *that satisfies* $|\{i \in S : i \le x\}| = \mathcal{O}(x/\log x)$*, we have that*

$$\lim_{N\to\infty} \frac{|\{i \le N : i \in f^{-1}(S)|}{N} = 0.$$

Then x_f *is continued fraction normal.*

For the proof, Vandehey uses metrical results to get asymptotics on how many rationals are m-good. In turn, these asymptotics imply conditions that determine whether the constructed number x_f is normal. This is an extension of the results in [1], where they used the Pointwise Ergodic Theorem when showing that nearly all rationals are m-good. Although this approach is standard in ergodic theory, it makes it unclear what the rate of convergence is [35]. By determining the aforementioned asymptotics, Vandehey can prove several new, concrete constructions of continued fraction normal numbers. One of the constructions for instance, considers the subsequence of rationals that have integer numerators and prime denominators [35, Theorem 1.4]. As such, the constructions from Vandehey and Adler, Keane and Smorodinsky are the only known concrete constructions for numbers that are continued fraction normal. In this thesis, we introduce a new type of normality for continued fractions and provide concrete constructions. Therefore, we have mostly focussed on - concrete - construction results. However, other type of results exist as well. We discuss some of these in the next section.

3.3.0 *Construction of μ-normal numbers, other results and final remarks*

Several types of normality results have been obtained since Borel introduced the notion of normal numbers in 1909. Each of the previous two sections discussed only one specific type of normality and focussed on results related to the construction of normal numbers. Next, we introduce a construction of a generalised form of normality and mention a few other type of results. We focus on results that are related to those in the previous sections and shed some light on others. We do not treat results in detail and do not cover the full variety of results. The goal is to provide some further historical background as well as perspective on the normality of numbers.

In 2016, Madrisch and Mance published a paper in which they provide a construction for a generalised type of normality [24]. Using symbolic dynamics, they generalise the notions of digits, blocks, number representations, shifts, concatenation and normality. This is followed by a construction of a sequence whose symbols (e.g. digits of the continued fraction) are distributed according to the invariant probability measure μ that is chosen. However, μ does not have to be the measure of maximal entropy (e.g. in the case of continued fraction expansions, μ does not have to be the Gauss measure). Let ω denote the infinite sequence of symbols that represents a number and let b denote any finite concatenation of symbols. Furthermore, define $G_n(\omega, b)$ as the number of occurrences of b in the first n symbols of ω. Then ω is μ-normal if

$$\lim_{n\to\infty} \frac{G_n(\omega, b)}{n} = \mu(b).$$

Thus, a number is μ-normal if the desired frequency of occurrences of any finite combination of symbols b, is specified by the measure that μ assigns to b. The method that is used in the construction is similar to the that of Copeland and Erdös [10] and also resembles that of Postnikov [29]. The authors construct an infinite sequence, such that each element of the sequence is approximately μ-normal. In order to do so, they take a sequence of measures $(\nu_i)_{i\geq 1}$ that converge in distribution to μ. Then, by using an analogue to Besicovitch's (ϵ, k)-normality, called (ϵ_i, k_i, ν_i)-normality, they construct an infinite sequence $(\omega_i)_{i\geq 1}$, where each ω_i is a finite concatenation of symbols that is (ϵ_i, k_i, ν_i)-normal. In other words, they construct an infinite sequence where each element is a better and better approximation of a μ-normal sequence. The μ-normal sequence is then obtained by concatenating a number of copies of ω_1, followed by more copies of ω_2, followed by even more copies of ω_3, and so on[1]. The chosen structure is necessary to guarantee that the construction works for a large class of different numeration systems. This is due to the fact that it allows them to control convergence by concatenating more and more copies of ω_n's, of which the distribution of symbols is closer and closer to μ as n tends to infinity. Although the authors apply the construction to different numeration systems such as b-adic expansions and continued fractions, they do not provide a concrete constructed number. However, up to our knowledge, the construction of a μ-normal number is the most generalised construction so far.

The aforementioned construction results are not the only results on normal numbers. Several other types of normality and existence results have been published since Borel introduced the notion of normality in 1909. Other results include normality for β-expansions, which are similar to normality for b-adic expansions. The difference is that b is a positive integer and β is allowed to be any positive real. Another normality result is given by Vandehey [33], who proved the theoretical existence of numbers that are both continued fraction normal and absolutely abnormal; not normal to any integer base. The proof, however, is conditional on the Generalized Riemann Hypothesis. Furthermore, Becher and Yuhjtman [2] provide an algorithm that proves the existence of numbers that are both continued fraction normal and absolutely normal; normal to every integer base. The key idea in their proof is to construct a sequence of nested intervals that satisfy certain conditions. Most of these conditions are related to discrepancy in the sense that they ensure an arbitrary small bound on the discrepancy of the numbers in that interval. The algorithm then ensures normality of the number that is obtained by taking the intersection

[1] See review of Madritsch and Mance [24].

of all these -sequences of nested- intervals. A similar approach is used by Madritsch, Scheerer and Tichy [25], who prove the existence of a number that is absolutely Pisot normal. That is, the constructed number is normal to each base from a given sequence of Pisot numbers, which are algebraic numbers with some special property. However, all known examples of (computable) absolutely normal numbers are given in the form of an algorithm [25].

Summarising, there is a wide variety of results on the normality of numbers. The results vary from concrete constructions to existence results based on algorithms. In order to prove the results, most authors have used techniques and concepts that are similar to those in the proof of Copeland and Erdös [10], yet other approaches exist as well. Moreover, concrete examples of normal numbers exist for different types of normality. The number of concrete examples, however, is small; especially when looking at the fact that almost all numbers are absolutely normal. The reason that it is so hard to exhibit normal numbers comes from the fact that it is practically impossible to consider an infinite sequence of digits. Moreover, the only way to consider the distribution of an infinite sequence of digits is to find a pattern in the behaviour of the sequence. The behaviour of an irrational numbers' number expansion however, generally does not exhibit a pattern. Therefore, people have constructed irrational numbers by defining a pattern, which is used to construct an irrational number. In turn, this pattern allows one to understand how the sequence behaves as it tends to infinity and therefore a distribution can be determined. In most types of normality, the desired distribution is the natural distribution, which comes from the measure of maximal entropy. As entropy is a measure of randomness, it follows that the number expansion of this normal number is (completely) random in some sense. Constructed normal numbers are such that the number expansion satisfies the distribution property that is associated with the corresponding type of normality. However, in a lot of cases, the expansion is not random at all. In other words, the behaviour of concrete constructed numbers is, in most cases, predictable. This makes sense, because it is this predictability - in the form of a pattern - that allows one to prove normality. To make a distinction between normal numbers that exhibit such a predictable pattern and those who do not, Adrian Belshaw introduced the concept of *strong normality* [3]. The current definition is only applicable to b-adic expansions and is yet to be generalized to other number expansions. In the case of b-adic expansions, Belshaw and Borwein [4] show that strongly normal numbers are normal and that the Champernowne number fails to be strongly normal. Furthermore, they prove that almost all numbers are strongly normal to all integer bases. More results related to strong normality can be found in [4] and [8]. This concludes the historical framework on normal numbers. Next, we introduce a new type of normality: *Minkowski normality.*

Chapter 4

A Minkowski normal number

Here, we introduce Minkowski normality and provide a concrete construction. That is, we introduce a type of normality for the continued fraction expansion that is related to the Minkowski question mark measure. In particular, we construct a number $\mathcal{K}$, see (4.10), of which the partial quotients are distributed according to the Minkowski question mark. This forms our main theorem, which is stated in Theorem 4.13. The crucial factor in determining the limiting distribution of this constructed number, is the ordering that is chosen. In the case of Adler, Keane and Smorodinsky [1], the ordering leads to normality with respect to the Gauss measure. Hence, the constructed number is continued fraction normal in the sense of Definition 3.11. In this chapter we consider the ordering that results from the so-called *Kepler tree*. We use this ordering to construct a number whose partial quotients are distributed according to the Minkowski question mark measure. We define this type of normality as follows.

Definition 4.1 (Minkowski normal number)**.** We say that $x = [a_1, a_2, a_3, \cdots] \in [0, 1)$ is *Minkowski normal*, if for any $k \geq 1$ and any block $d = d_1, d_2, \cdots, d_k$, with $d_i \in \mathbb{N}$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\Delta(d)}(\mathcal{T}^i x) = 2^{-(d_1+d_2+\cdots+d_k)}. \tag{4.1}$$

Theorem 4.2. *$\mu_?$ almost every number in $[0, 1)$ is Minkowski normal.*

Proof. Let $x \in [0, 1)$ and consider the ergodic system $(X, \mathcal{B}, \mu_?, \mathcal{T})$. Then for any $k \geq 1$ and any block $d = d_1, d_2, \cdots d_k$, $d_i \in \mathbb{N}$, it follows from Theorem 2.6 that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\Delta(d)}(\mathcal{T}^i x) = \mu_?(\Delta(d)) = 2^{-(d_1+d_2+\cdots+d_k)} \qquad a.e.$$

□

Constructing a number and proving that it is Minkowski normal is the main objective of this thesis. The corresponding theorem and proof are stated in section 4.5. First however, we discuss results from Kessebömer and Stratmann, which have been the initial motivation for this research. Then, we discuss the construction and introduce tools that are used in the proof. After proving Minkowski normality of the constructed number, we conclude the chapter with a numerical experiment and final remarks.

4.1.0 *Kessebömer, Stratmann and the Stern-Brocot sequence*

Kessebömer and Stratmann have published two papers that sparked the idea for this research. In 2012 they published a paper concerning the distribution of the (weighted) *Farey* and *even Stern-Brocot sequence* [23]. This paper also refers to a paper of the two authors that was published four years earlier, which studies properties of the Minkowski question mark [22]. We next introduce the two sequences and briefly discuss results from Kessebömer and Stratmann. Finally, we conclude this section by relating the two papers and this research.

The sequences that Kessebömer and Stratmann discuss are sequences of sets containing fractions between 0 and 1. The Farey $(\mathcal{F}_i)_{i\geq 1}$ sequence can be characterised as the sequence of sets such that the n-th set contains all irreducible fractions that have a denominator lower or equal to n. That is,

$$\mathcal{F}_n := \{p/q : 0 < p \leq q \leq n, \gcd(p,q) = 1\}, \qquad n \geq 1. \tag{4.2}$$

The first four sets are

$$\begin{aligned}
\mathcal{F}_1 &= \{\tfrac{0}{1}, && && && && && && \tfrac{1}{1}\},\\
\mathcal{F}_2 &= \{\tfrac{0}{1}, && && && \tfrac{1}{2}, && && && \tfrac{1}{1}\},\\
\mathcal{F}_3 &= \{\tfrac{0}{1}, && && \tfrac{1}{3}, && \tfrac{1}{2}, && \tfrac{2}{3}, && && \tfrac{1}{1}\},\\
\mathcal{F}_4 &= \{\tfrac{0}{1}, && \tfrac{1}{4}, && \tfrac{1}{3}, && \tfrac{1}{2}, && \tfrac{2}{3}, && \tfrac{3}{4}, && \tfrac{1}{1}\}.
\end{aligned}$$

The definition for the even Stern-Brocot sequence $(\mathcal{S}_i)_{i\geq 1}$ is more technical. For this sequence of sets, the n-th set is given by

$$\mathcal{S}_n := \{s_{n,2k}/t_{n,2k} : k = 1, 2, \ldots, 2^{n-1}\}, \qquad n \geq 1, \tag{4.3}$$

where $s_{0,1} := 0$, $s_{0,2} := t_{0,1} := t_{0,2} = 1$. Furthermore, for $\alpha \in \{s, t\}$, we define

$$\begin{aligned}
\alpha_{n+1,2k-1} &:= \alpha_{n,k}, && k = 1, 2, \ldots, 2^n + 1,\\
\alpha_{n+1,2k} &:= \alpha_{n,k} + \alpha_{n,k+1}, && k = 1, 2, \ldots, 2^n.
\end{aligned}$$

Using the latter recurrence, we see that the fractions in $\mathcal{S}_{n+1}$ are obtained by taking mediants

$$\frac{s_{n+1,2k}}{t_{n+1,2k}} = \frac{s_{n,k} + s_{n,k+1}}{t_{n,k} + t_{n,k+1}}.$$

Moreover, fixing n, we see that the rationals in $\mathcal{S}_{n+1}$ are obtained by taking mediants from two consecutive rationals in the sequence $(s_{n,k}/t_{n,k})_{k=1}^{2^n+1}$. For reasons of convenience, we define $\mathcal{S}_0 = \{\frac{0}{1}, \frac{1}{1}\}$. We then find that the first four sets of the sequence $(\mathcal{S}_i)_{i\geq 0}$ are

$$\begin{aligned}
\mathcal{S}_0 =&\{\tfrac{0}{1}, && && && && && && && && \tfrac{1}{1}\},\\
\mathcal{S}_1 =&\{ && && && && \tfrac{1}{2} && && && && \},\\
\mathcal{S}_2 =&\{ && && \tfrac{1}{3}, && && && && \tfrac{2}{3} && && \},\\
\mathcal{S}_3 =&\{ && \tfrac{1}{4}, && && \tfrac{2}{5}, && && \tfrac{3}{5}, && && \tfrac{3}{4} && \}.
\end{aligned}$$

This sequence is called the **even** Stern-Brocot sequence, because it is a subsequence of the *Stern-Brocot sequence*. The latter sequence, $(\bar{\mathcal{S}_i})_{i\geq 0}$, is defined similar to $(\mathcal{S}_i)_{i\geq 1}$. They use the same recurrence relations and, instead of (4.3), the n-th set of the Stern-Brocot sequence is given by

$$\bar{\mathcal{S}_n} := \{s_{n,k}/t_{n,k} : k = 1, 2, \ldots, 2^n\}, \qquad n \geq 0. \tag{4.4}$$

Moreover, this sequence corresponds to the level sets of the *Stern-Brocot tree*. This tree is formed by starting with $\frac{0}{1}$ and $\frac{1}{0}$ and inserting the median between two adjacent fractions, see Figure 4.1. If we denote the n-th level set of the Stern-Brocot tree by $\mathcal{L}_n$, then we have

$$\mathcal{S}_n = \{\tfrac{p}{q} \in \mathcal{L}_n : p < q\} \quad \text{and} \quad \mathcal{F}_n = \{\tfrac{p}{q} \in \cup_{i=1}^{n}\mathcal{L}_i : p < q \leq n\} \cup \{\tfrac{0}{1}, \tfrac{1}{1}\}, \qquad n \geq 1.$$

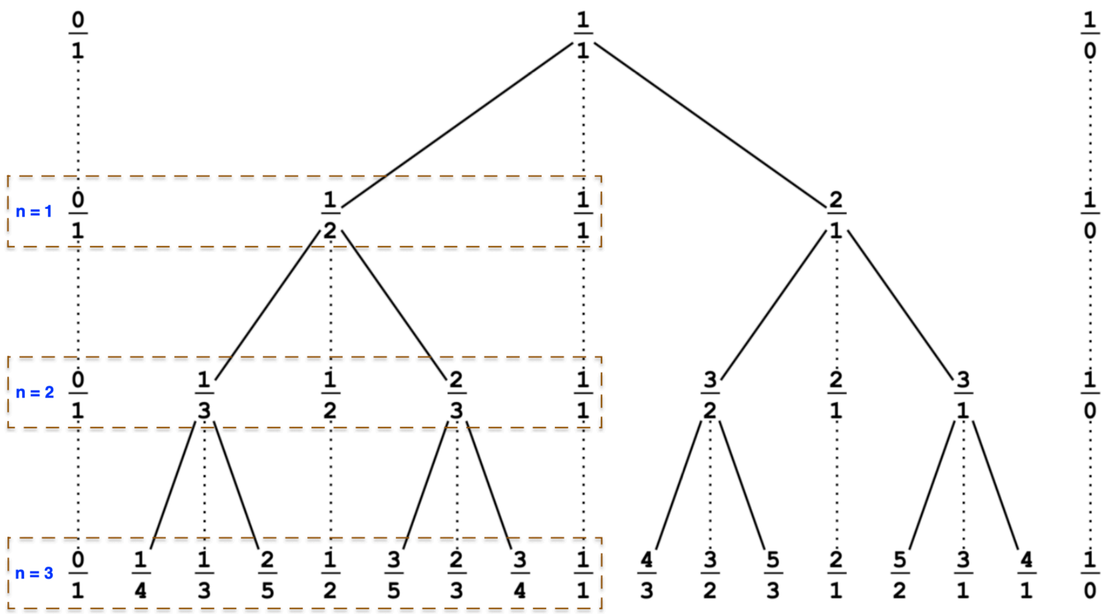


Figure 4.1: The Stern-Brocot tree.

The relevant fractions correspond to the left half of the Stern-Brocot tree and are boxed in the dashed brown boxes in Figure 4.1. In words, the n-th set of the Farey sequence can be obtained by taking all fractions in the left half of the Stern-Brocot tree, from the first up to the n-th level, leaving out the fractions with denominator greater than n. Furthermore, the n-th set of the even Stern-Brocot sequence corresponds to the proper fractions in n-th level set the Stern-Brocot tree. Moreover, the even Stern-Brocot sequence corresponds to the level sets of the subtree that has $1/2$ as a root. Remarkably, this subtree is also referred to as the *Farey tree* [7]. We discuss this tree in more detail in chapter 5. For now, we conclude that the Farey sequence and even Stern-Brocot sequence are closely related.

The main result of the 2012 paper is a "dichotomy between uniform distributions of the

Stern-Brocot and the Farey sequence". More specifically, they prove the following.

Theorem 4.3. *[23, Theorem 1.1] For the even Stern-Brocot sequence we have*

$$\log(n^2) \sum_{p/q \in \mathcal{S}_n} \frac{1}{q^2} \delta_{p/q} \xrightarrow{d} \lambda, \tag{4.5}$$

and for the Farey sequence

$$\frac{\zeta(2)}{\log(n)} \sum_{p/q \in \mathcal{F}_n} \frac{1}{q^2} \delta_{p/q} \xrightarrow{d} \lambda. \tag{4.6}$$

The theorem thus states that the weighted sum of Dirac measures on the sequence converges weakly to the Lebesgue measure. That is, with the appropriate canonical weights, the sequences eventually distribute rationals uniformly. This convergence is complimentary to earlier results, which proved weak convergence of the uniformly weighted sequence. These results show that

$$\frac{1}{\#\mathcal{F}_n} \sum_{p/q \in \mathcal{F}_n} \delta_{p/q} \xrightarrow{d} \lambda \tag{4.7}$$

and

$$\frac{1}{\#\mathcal{S}_n} \sum_{p/q \in \mathcal{S}_n} \delta_{p/q} \xrightarrow{d} \mu_?, \tag{4.8}$$

where $\#A$ denotes the cardinality of the set A. The latter convergence result is an immediate consequence of [22, Proposition 3.1]. The convergence in (4.8) tells us that the even Stern-Brocot sequence approximately distributes the rationals according to the Minkowski question mark measure. It is this convergence, that sparked the idea for this research. Moreover, in our construction we consider a similar, but not equivalent, ordering of the rationals. One of the differences is that, contrary to the construction, the convergence in (4.8) does not use an explicit ordering of the rationals in the set. Furthermore, weak convergence of the Stern-Brocot sequence implicitly uses a rationals full continued fraction expansion, whereas the normality condition as stated in Definition 4.1 requires blocks of arbitrary length to be distributed according to the Minkowski question mark measure. These arbitrary blocks can also be a part of a rationals continued fraction expansion. We further elaborate on the differences and similarities in chapter 5, where we reflect on the construction and extend our results. Among others, we show that the number that is obtained by concatenating the continued fraction expansions of the rationals in the even Stern-Brocot sequence is Minkowski normal as well. First, we elaborate on the construction and prove that resulting number, see (4.10), is Minkowski normal.

4.2.0 *Kepler and the construction*

The construction of a Minkowski normal number is based on the Kepler tree. As such, we introduce the tree and some of its relevant properties. We explicitly construct an irrational number by defining an infinite sequence of rationals of which we concatenate the continued fraction expansions. We then construct a binary analogue, which plays a key role in proving normality of the constructed number.

The first part of the Kepler tree is found in Johannes Kepler's magnum opus, a book containing his most important work. In this work he discusses harmonic divisions of strings and displays the first few levels of the tree, starting from $1/1$. See [21, p. 163] for an English translation. Though Johannes Kepler starts from $1/1$, the tree starts from $1/2$ and then uses the rule

$$\begin{tikzcd}[column sep=small] & p/q \arrow[dl, dash] \arrow[dr, dash] & \\ p/(p+q) & & q/(p+q) \end{tikzcd}.$$

Also, as rationals can be represented by finite continued fractions and vice versa [27], we introduce the following lemma. Note that we assume continued fractions to be written in their reduced form. That is, the last digit of the expansion is greater or equal than 2.

Lemma 4.4. *Let p/q be an arbitrary rational and let $[a_1, a_2, \cdots, a_n]$ denote the corresponding continued fraction expansion. Then the Kepler rule can be represented as*

$$\begin{tikzcd}[column sep=small] & {[a_1, a_2, \cdots, a_n]} \arrow[dl, dash] \arrow[dr, dash] & \\ {[(a_1+1), a_2, \cdots, a_n]} & & {[1, a_1, a_2, \cdots, a_n]} \end{tikzcd}.$$

Proof. The equivalence can be seen from

$$\begin{aligned} \frac{p}{p+q} = \frac{1}{1+\frac{q}{p}} = \frac{1}{1+\cfrac{1}{\frac{p}{q}}} = \frac{1}{1+\frac{1}{[a_1,a_2,\cdots,a_n]}} \\ = [a_1+1, a_2, \cdots, a_n] \end{aligned}$$

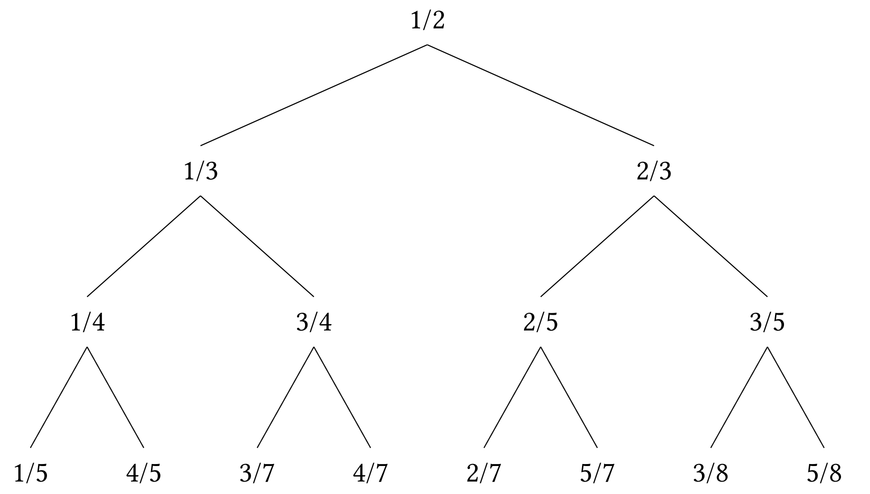
and

$$\begin{aligned} \frac{q}{p+q} = \frac{1}{\frac{p}{q}+1} = \frac{1}{1+[a_1,a_2,\cdots,a_n]} \\ = [1, a_1, a_2, \cdots, a_n]. \end{aligned}$$

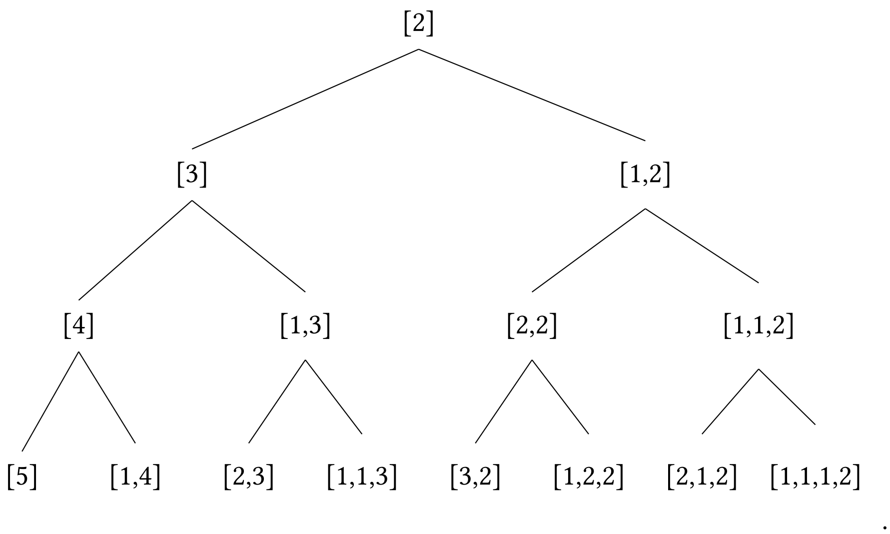
□

Note that a left move increases the first digit in the continued fraction by one and does not alter the total number of digits in the continued fraction. A right move however, inserts a 1 as a first digit and thus increases the length of the continued fraction by one. This also means that a left move does not preserve the block of digits that form the continued fraction of the mother node, whereas a right move does preserve the block. Lastly, note that both moves increase the sum of the digits of the continued fraction expansion by one.

The tree thus starts with $1/2$ at the root. We refer to the root as level 0 and generate the subsequent levels using the Kepler rule. The first four levels of the tree are displayed in Figure 4.2.



(a) Ordinary Kepler tree.



(b) Continued fraction Kepler tree.

Figure 4.2: The first 4 levels of the Kepler tree, expressed in (a) rationals and (b) the corresponding continued fraction expansions.

Subsequently, we concatenate the continued fractions of the rationals in the tree going top-down, left-right. The ordering of the rationals that result from this procedure is

$$\frac{1}{2}, \quad \frac{1}{3}, \frac{2}{3}, \quad \frac{1}{4}, \frac{3}{4}, \frac{2}{5}, \frac{3}{5}, \quad \frac{1}{5}, \cdots \tag{4.9}$$

and we denote the corresponding sequence of blocks by $(k_i)_{i\geq 1}$. That is, k_i denotes the block of digits that form the continued fraction expansion of the i-th rational in (4.9). The above sequence can then be given by

$$[k_1], \quad [k_2], [k_3], \quad [k_4], [k_5], [k_6], [k_7], \quad [k_8], \cdots .$$

By concatenating the blocks $(k_i)_{i\geq 1}$, we construct an infinite continued fraction, which is a unique irrational number by Proposition 2.14. The resulting number is given by:

$$\mathcal{K} := [2, \quad 3, \quad 1, 2, \quad 4, \quad 1, 3, \quad 2, 2, \quad 1, 1, 2, \quad 5, \quad \cdots] \approx 0.44031. \tag{4.10}$$

The key idea in proving normality of $\mathcal{K}$ is that we can use binary codes to identify arbitrary blocks in the Kepler tree. In order to do this, we first construct a binary analogue for the (continued fraction) Kepler tree.

4.3.0 *Retracing paths and constructing a binary analogue*

Before we prove that $\mathcal{K}$ is Minkowski normal, we construct an analogue to the previous section. The underlying idea is that, given an arbitrary rational number, we can retrace its position in the Kepler tree. Moreover, we can retrace the path between the root and the rational. Using these paths, we associate a binary code to each rational in the Kepler tree. Consequently, we replace the rationals in the Kepler tree by their corresponding binary code and hence construct another tree. Then, similar to the construction of $\mathcal{K}$, we construct another number, which turns out to be a well-known example of a number that is normal in base 2.

Retracing paths is possible due to the result in Lemma 4.4. More specifically, it is due to the fact that a left move in the Kepler tree increases the first digit of the continued fraction by 1 and a right move inserts the digit 1 as the first digit of the continued fraction. We retrace the path of a rational as follows. Let $[a_1, a_2, \cdots, a_n]$ denote the continued fraction of an arbitrary rational p/q in the Kepler tree. Then by going $(a_1 - 1)$ steps from the left up, we end up at the rational $[1, a_2, \cdots, a_n]$. Subsequently, going from the right up we end at $[a_2, a_3, \cdots, a_n]$. By repeating this proces for $a_2, a_3, \ldots, a_{n-1}$ and a_n we find the path to the root. This is summarised schematically in Figure 4.3. Note that it takes $(a_n - 2)$ steps from the left up to get to the root from $[a_n]$. This is due to the fact that we end up at the digit 2 instead of 1. If we summarise the upward path symbolically by writing L for a left move and R for a right move, we find that the upward path is given by $L^{a_1-1}RL^{a_2-1}R\cdots L^{a_n-2}$. By reversing this path we obtain the downward path; the path from the root to the rational. Hence, the downward path that corresponds to the rational with continued fraction expansion $[a_1, a_2, \cdots, a_n]$ is given by $L^{a_n-2}\cdots RL^{a_2-1}RL^{a_1-1}$. Furthermore, the total number of steps in this path is

$a_1 + a_2 + \cdots + a_n - 2$, which also corresponds to level in which the block $a_1, a_2, \cdots, a_n$ occurs for the first time.

Example 4.5. Consider the rational $4/7$, which is represented by the continued fraction $[1, 1, 3]$, see Figure 4.2(b). In order to find the upward path, we first make a right step upward, ending up at $[1, 3]$. Another right step upwards takes us to $[3]$, after which a left step upward takes us to the root. The upward path then becomes R^2L. Hence the downward path is given by LR^2. Thus starting from the root, $[2]$, a left move first increases the first digit of the continued fraction. This results in the continued fraction $[3]$. The two subsequent right moves, consecutively insert a 1 at the start of the continued fraction. This results in the continued fraction $[1, 1, 3] = 4/7$. Also, note that the rational occurs in level $3 = 1 + 1 + 3 - 2$.

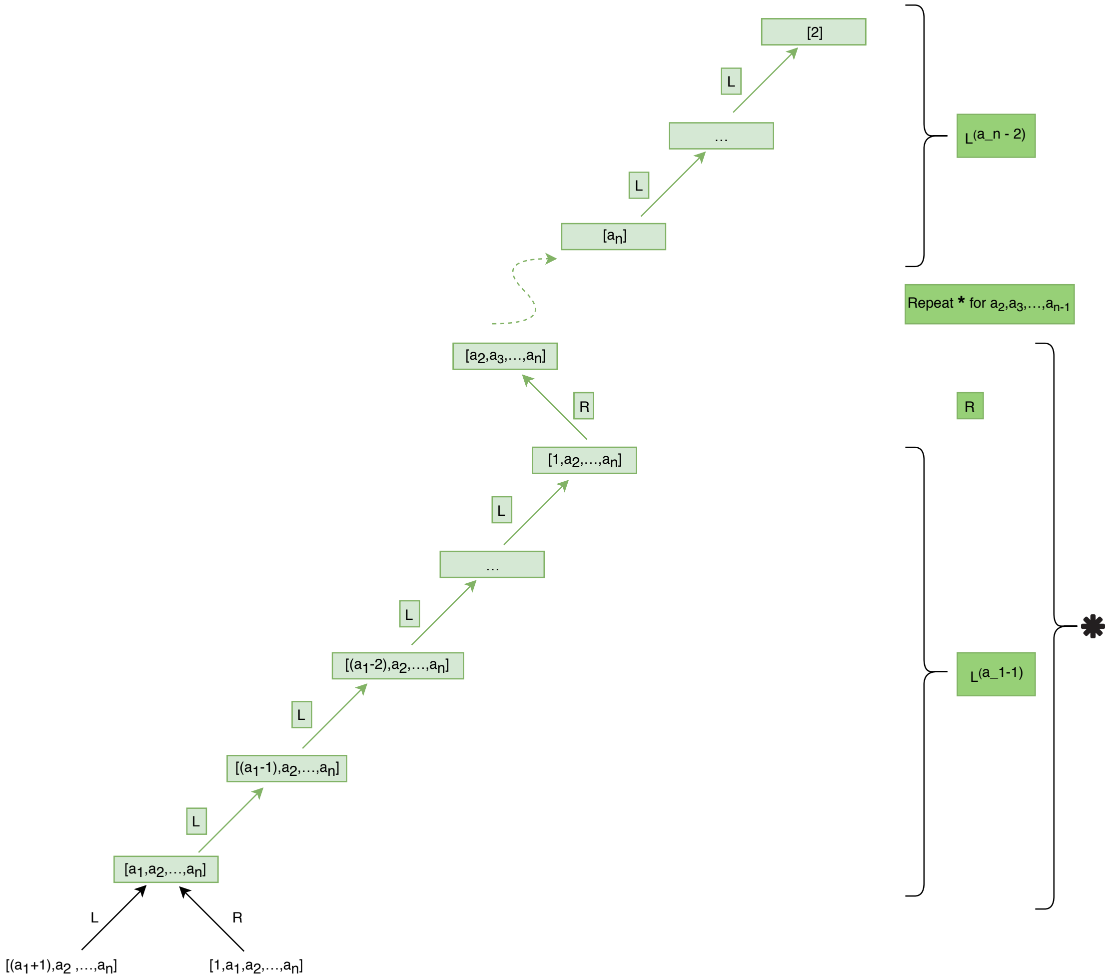


Figure 4.3: Retracing the upward path in the Kepler tree for an arbitrary rational.

Next, we associate each rational in the Kepler tree with a binary code by retracing its downward path and then applying the substitution $\{L \mapsto 0, R \mapsto 1\}$. The root is associated to the empty word. Furthermore, let $[a_1, a_2, \cdots, a_n]$ denote the continued fraction expansion of an arbitrary rational p/q. Then we associate a binary code to p/q as follows

$$\begin{aligned} p/q \xleftrightarrow{\text{cfe}} [a_1, a_2, \cdots, a_n] &\xleftrightarrow{\text{upward path}} L^{a_1-1}RL^{a_2-1}R\cdots L^{a_n-2} \\ &\xleftrightarrow{\text{downward path}} L^{a_n-2}\cdots RL^{a_2-1}RL^{a_1-1} \\ &\xleftrightarrow{\text{binary code}} 0^{a_n-2}\cdots 10^{a_2-1}10^{a_1-1}. \end{aligned}$$

Subsequently, we construct another binary tree by replacing the rationals in the Kepler tree by their corresponding binary code, see Figure 4.4. We refer to this tree as the *binary Kepler tree.*

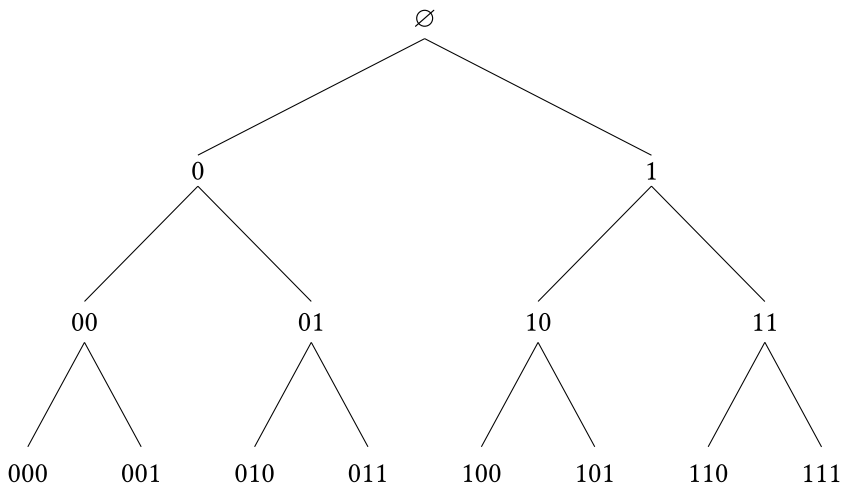
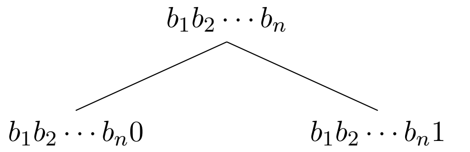


Figure 4.4: The first 4 levels of the binary Kepler tree.

Let $b = b_1b_2\cdots b_n$ be an arbitrary binary code, $b_i \in \{0, 1\}$, $1 \le i \le n$. Then the rule in the binary Kepler tree is given by



Note that a left move in the binary Kepler tree appends the digit 0 at the end of the binary code and a right move appends a 1. Thus once a block occurs within the binary Kepler tree, it is preserved forever. This is different from the Kepler tree, which only preserves blocks by making a right move. Furthermore, note that moves in the Kepler tree correspond to changes at the start of a continued fraction expansion, whereas moves in the binary tree correspond to changes at the end of a binary code.

Lastly, analogous to the construction of $\mathcal{K}$, we construct an infinite binary code by concatenating the binary codes going top-down, left-right. The ordering of binary codes that is obtained through this is

$$\varnothing, 0, 1, 00, 01, 10, 11, 000, \cdots . \tag{4.11}$$

and we denote the corresponding sequence of binary codes by $(c_i)_{i\geq 1}$. That is, c_i denotes the i-th binary code in (4.11). By concatenating the blocks $(c_i)_{i\geq 1}$, we construct an infinite sequence $c_1c_2c_3\cdots$ and define the dyadic Champernowne number

$$\mathcal{C}_2 := 0.c_1c_2c_3\cdots = 0.\ 0\ 1\ 00\ 01\ 10\ 11\ 000\ \cdots , \tag{4.12}$$

which is known to be normal in base 2. This and other properties of $\mathcal{C}_2$ can for instance be found in [14] or [31].

Theorem 4.6. *The number $\mathcal{C}_2$ is normal in base 2.*

Suppose that $b = b_1b_2\cdots b_l$ is an arbitrary binary code of length l. The base 2 normality of $\mathcal{C}_2$ then implies that the asymptotic frequency of b in $\mathcal{C}_2$ equals 2^{-l}, see Definition 3.1. Thus if we can identify binary blocks that correspond to occurrences of arbitrary blocks in the Kepler tree, we can use the normality of $\mathcal{C}_2$ to prove normality of $\mathcal{K}$. This is the topic of the next section.

4.4.0 *Relating the constructions*

In this section we relate the Kepler tree and the binary analogue. Moreover, we discuss how blocks are formed and preserved in the Kepler tree and how we can use binary codes to identify them. The results in this section are the basis for the proof that $\mathcal{K}$ is Minkowski normal.

Lemma 4.7. *There exists a one-to-one correspondence between the Kepler tree and the binary Kepler tree.*

Proof. Let p/q be an arbitrary rational in the Kepler tree and suppose that $[a_1, a_2, \cdots, a_n]$ is the continued fraction expansion that corresponds to p/q. Then the correspondence between the trees is given by $p/q \longleftrightarrow 0^{a_n-2}\cdots 10^{a_2-1}10^{a_1-1}$. In other words, there exists a one-to-one correspondence between $(k_i)_{i\geq 1}$ and $(c_i)_{i\geq 1}$, which is given by

$$k_i \mapsto c_i$$

for all $i \geq 1$.

$i:$	1	2	3	4	5	6	7	$8\cdots$
$k_i:$	2	3	$1,2$	4	$1,3$	$2,2$	$1,1,2$	$5\cdots$
	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow\cdots$
$c_i:$	$\varnothing$	0	1	00	01	10	11	$000\cdots$.

The correspondence being one-to-one follows from uniqueness of the paths from which the binary code results. $\square$

The correspondence between the trees provides a lot of information. More specifically, the binary code that is associated to a rational contains a lot of information. It gives the continued fraction expansion of the rational that it represents and its exact location within the tree. Namely, it gives the level in which the rational occurs and the position within that level. The level is given by the total number of 0's and 1's in its binary code and its position within the level can be read from the ordering of the 0's and 1's. The following lemma is an immediate consequence of the binary analogue and the concept of retracing paths in the tree.

Lemma 4.8. *There exists a unique path between the root of the Kepler tree that starts at $1/2$ and any arbitrary rational p/q. If we denote p/q by its continued fraction expansion $[a_1, a_2, \cdots, a_n]$, then the corresponding path is*

$$L^{a_n-2} \cdots RL^{a_2-1}RL^{a_1-1}, \tag{4.13}$$

which corresponds to the binary code

$$0^{a_n-2} \cdots 10^{a_2-1}10^{a_1-1}. \tag{4.14}$$

This path consists of $a_1 + a_2 + \cdots + a_n - 2$ moves, which also corresponds to the level in which the rational occurs for the first and only time.

Apart from providing information about the occurrence of rationals, the concept of retracing paths also tells us how blocks of the form $d = d_1, d_2, \cdots, d_k$ are formed by the Kepler tree, how these blocks are preserved and how we can identify them using binary codes. This is the key takeaway of the next lemma.

Lemma 4.9. *Let $[a_1, a_2, \cdots, a_n]$ denote the continued fraction expansion of an arbitrary rational p/q in the Kepler tree and let $d = d_1, d_2, \cdots, d_k$ be an arbitrary block of length k. Then there exists a unique subpath from p/q to the rational r/s that corresponds to the continued fraction $[d_1, d_2, \cdots, d_k, a_1, a_2, \cdots, a_n]$. This subpath consists of $d_1 + d_2 + \cdots + d_k$ moves and is given by*

$$RL^{d_k-1} \cdots RL^{d_2-1}RL^{d_1-1}, \tag{4.15}$$

which corresponds to the binary code

$$10^{d_k-1} \cdots 10^{d_2-1}10^{d_1-1}. \tag{4.16}$$

Proof. By Lemma 4.8, there exists a unique path to p/q that is given by

$$L^{a_n-2} \cdots RL^{a_2-1}RL^{a_1-1}.$$

Similarly, there exists a unique path to the rational $[d_1, d_2, \cdots, d_k, a_1, a_2, \cdots, a_n]$. By (4.13), this path is given by

$$\boldsymbol{L^{a_n-2} \cdots RL^{a_2-1}RL^{a_1-1}}RL^{d_k-1} \cdots RL^{d_2-1}RL^{d_1-1}.$$

Considering the latter path, we see that it passes through the rational p/q, of which the path is marked in bold. As this path and that to p/q are unique, we conclude that there exists a unique subpath from p/q to $[d_1, d_2, \cdots, d_k, a_1, a_2, \cdots, a_n]$ that is given by

$$RL^{d_k-1} \cdots RL^{d_2-1}RL^{d_1-1},$$

In order to see that this path consists of $d_1 + d_2 + \cdots + d_k$ steps, we can do two things. One is to count the number of L's and R's in the subpath. The other way to see this, is from the fact that p/q uniquely occurs in level $a_1 + a_2 + \cdots + a_n - 2$ and that $[d_1, d_2, \cdots, d_k, a_1, a_2, \cdots, a_n]$ occurs uniquely in level $d_1 + d_2 + \cdots + d_k + a_1 + a_2 + \cdots + a_n - 2$. Hence the number of steps is given by the difference between these levels

$$(d_1 + d_2 + \cdots + d_k + a_1 + a_2 + \cdots + a_n - 2) - (a_1 + a_2 + \cdots + a_n - 2) = d_1 + \cdots + d_k.$$

$\square$

In other words, Lemma 4.9 provides information about blocks occurring at the start of a continued fraction expansion and how we can identify such occurrences using binary codes. Moreover, the lemma implicitly describes how blocks occur in the middle of a continued fraction. To see this, consider an arbitrary block $c = c_1, c_2 \cdots, c_m$ and apply Lemma 4.9 to the concatenation of c and d: $c_1, c_2, \cdots, c_m, d_1, d_2 \cdots, d_k$. Then d occurs in the middle of the continued fraction expansion of the rational that corresponds to $[c_1, c_2, \cdots, c_m, d_1, d_2, \cdots, d_k, a_1, a_2, \cdots, a_n]$. The path that is associated to this rational is

$$L^{a_n-2} \cdots RL^{a_2-1}RL^{a_1-1}\boldsymbol{RL^{d_k-1} \cdots RL^{d_2-1}RL^{d_1-1}R}L^{c_m-1} \cdots RL^{c_2-1}RL^{c_1-1},$$

with in bold the part of the path that corresponds to forming d in the middle. Note that this is different from the subpath in (4.15). Namely, there is an extra R at the end of the bold part. This right move is necessary to preserve the block and causes it to occur in the middle. Hence we conclude that the binary code that is associated to this type of occurrence is given by

$$10^{d_k-1} \cdots 10^{d_2-1}10^{d_1-1}1. \tag{4.17}$$

The lemmas that are presented in this section provide information about the way that blocks are formed and occur in the Kepler tree and how we can use binary codes to identify them. These lemmas and the correspondence between the Kepler tree and the binary Kepler tree are essential for proving that $\mathcal{K}$ is Minkowski normal. The latter is essential because of the fact that $\mathcal{C}_2$ is normal in base 2. Furthermore, it is important to note that the correspondence in Lemma 4.7 is a correspondence between rationals and binary codes. In other words, it tells us that each binary code corresponds to a rational and vice versa. In order to prove normality of $\mathcal{K}$, we have to consider occurrences of arbitrary blocks, which can be part of a rationals continued fraction expansion. Hence, when we use $\mathcal{C}_2$ to prove normality of $\mathcal{K}$, we need to identify binary codes that correctly correspond to occurrences of the (corresponding) arbitrary blocks in $\mathcal{K}$. Equations (4.14), (4.16) and (4.17) partially provide the answer. We complete the argument in the next section, where we prove that $\mathcal{K}$ is Minkowski normal.

4.5.0 *Proof of normality*

The lemmas that are presented in the previous section provide information on how to use binary codes to identify occurrences of arbitrary blocks in $\mathcal{K}$. In order to prove that $\mathcal{K}$ is Minkowski normal, we identify explicit binary codes that correspond to different types of occurrences of blocks in $\mathcal{K}$. Consequently, we use the base 2 normality of $\mathcal{C}_2$ to determine the frequency that corresponds to these type of occurrences. This approach, however, does not work for one type of occurrence. In order to deal with this exception, we show that the frequency of this type of occurrence tends to zero. We then state the main theorem of the thesis and prove that the frequency of the remaining type of occurrences is the desired frequency.

Apart from the exception, the different types of occurrences relate to the results in Lemma 4.8 and 4.9. We distinguish the following types of occurrences.

- The block occurs at the start of a continued fraction expansion of a rational;
- The block occurs in the middle of the continued fraction expansion of a rational;
- The block occurs at the end of the continued fraction expansion of a rational;
- The block occurs as a result of concatenating the continued fraction expansions of different rationals. We refer to this type of occurrences as *divided occurrences*.

The first three types are related to the lemmas and the latter type is the exception. To see why the latter is an exception, let $d = d_1, d_2, \cdots, d_k$ be an arbitrary block of length k, $d_i \in \mathbb{N}$. For d to occur divided, the elements of d should be split over two or more continued fraction expansions. For simplicity, suppose that the occurrence of d results from a concatenation of two rationals r_1 and r_2, where r_1 is the left neighbour of r_2. Let $[a_1, a_2, \cdots, a_n]$ and $[b_1, b_2, \cdots, b_m]$ denote the continued fraction expansions that correspond to r_1 and r_2, respectively. If d results from concatenation, then for some $j < k$, the block $d_1, d_2, \cdots, d_j$ occurs at the end of the continued fraction expansion of r_1 and $d_{j+1}, d_{j+2}, \cdots d_k$ occurs at the start of the continued fraction expansion of r_2. Moreover, we have the following identities

$$\begin{aligned} a_{n-j+i_1} &= d_{i_1}, \\ b_{i_2} &= d_{j+i_2}, \end{aligned}$$

for $1 \leq i_1 \leq j$ and $1 \leq i_2 \leq k-j$. Using these identities, we rewrite the binary codes that correspond to r_1 and r_2 as

$$\begin{aligned} r_1 \longleftrightarrow & 0^{a_n-2} \cdots 10^{a_{n-j+2}-1} 10^{a_{n-j+1}-1} 10^{a_{n-j}-1} \cdots 10^{a_2-1} 10^{a_1-1} \\ = & \mathbf{0^{d_j-2} \cdots 10^{d_2-1} 10^{d_1-1}} 10^{a_{n-j}-1} \cdots 10^{a_2-1} 10^{a_1-1} \end{aligned}$$

and

$$\begin{aligned} r_2 \longleftrightarrow & 0^{b_m-2} \cdots 10^{b_{k-j+1}-1} 10^{b_{k-j}-1} \cdots 10^{b_2-1} 10^{b_1-1} \\ = & 0^{b_m-2} \cdots 10^{b_{k-j+1}-1} \mathbf{10^{d_k-1} \cdots 10^{d_{j+2}-1} 10^{d_{j+1}-1}}, \end{aligned}$$

where the bold parts correspond to forming the digits that, together, compose d. When concatenating elements in the binary Kepler tree we concatenate the binary codes of r_1 and r_2 as follows

$$\mathbf{0^{d_j-2}\cdots 10^{d_2-1}10^{d_1-1}}10^{a_{n-j}-1}\cdots 10^{a_2-1}10^{a_1-1}0^{b_m-2}\cdots$$
$$10^{b_{k-j+1}-1}\mathbf{10^{d_k-1}\cdots 10^{d_{j+2}-1}10^{d_{j+1}-1}}.$$

Note that the parts of the binary code that correspond to digits of d will be separated and that the intermediate part may vary per divided occurrence. This makes it impossible to identify an explicit binary code that corresponds to divided occurrences. In order to circumvent this problem, we introduce two lemmas. The first lemma relates to the total number of possible occurrences of an arbitrary block in the Kepler tree and the second lemma states that the frequency of divided occurrences tends to zero.

Lemma 4.10. *The total number of digits in the l-th level of the Kepler tree is $(l+2)2^{-1}$, $l \geq 0$.*

Proof. For this proof we use the binomial identity, which states that

$$(x+y)^l = \sum_{i=0}^{l} \binom{l}{i} x^i y^{l-i}.$$

The l-th row of the Kepler tree consists of 2^l rationals. Each of these are formed by i left moves and $l-i$ right moves, where i varies between 0 and l. Each right move increases the number of digits by 1 and we start off with one digit at level 0. Therefore, the total number of digits in level l is given by

$$\begin{aligned}\sum_{i=0}^{l}(i+1)\binom{l}{i} &= \sum_{i=0}^{l}\binom{l}{i} + \sum_{i=0}^{l} i\binom{l}{i}\\ &= 2^l + l2^{l-1}\\ &= (l+2)2^{l-1}\end{aligned}$$

More specifically we use the binomial identity as follows. For the first sum we take x and y to be 1. This gives

$$\sum_{i=0}^{l}\binom{l}{i} = \sum_{i=0}^{l}\binom{l}{i}1^i1^{l-i} = (1+1)^l = 2^l.$$

For the second sum, we first set $y=1$,

$$(x+1)^l = \sum_{i=0}^{l}\binom{l}{i}x^i,$$

and then differentiate both sides with respect to x. This gives

$$l(x+1)^{l-1} = \sum_{i=0}^{l} i \binom{l}{i} x^{i-1}.$$

The desired result is then obtained by taking $x = 1$. $\square$

Corollary 4.11. Let $d = d_1, d_2, \cdots, d_k$ be an arbitrary block of length k. The number of possible occurrences of d in the l-th level of the Kepler tree is given by

$$(l+2)2^{l-1} - k + 1. \tag{4.18}$$

Using this corollary, we prove that the frequency of divided occurrences tends to zero.

Lemma 4.12. *Let $d = d_1, d_2, \cdots, d_k$ be an arbitrary block of length k, $d_i \in \mathbb{N}$. The asymptotic frequency of divided occurrences of d in $\mathcal{K}$ is equal to 0.*

We give a proof using techniques similar to that in Champernowne [9].

Proof. Each level l in the Kepler tree consists of 2^l rationals, which results in a total of $2^l - 1$ concatenations. Furthermore, d consists of k digits, which results in a maximum of $k-1$ positions where d can be divided. Therefore, the number of divided occurrences can be bounded from above by $k2^l$. We know from Corollary 4.11 that the total number of possible occurrences of d in the l-th level is given by $(l+2)2^{l-1} - k + 1$.

Next, suppose that the n-th digit of $\mathcal{K}$ occurs within the L-th level of the Kepler tree. The number of divided occurrences in the first n digits of $\mathcal{K}$ is then bounded from above by

$$\sum_{l=0}^{L-1} k2^l + \mathcal{O}(2^L) = k(2^L - 1) + \mathcal{O}(2^L).$$

Furthermore, we find that the total number of possible occurrences of d in the first n digits of $\mathcal{K}$ is

$$\sum_{l=0}^{L-1} (l+2)2^{l-1} - k + 1 + \mathcal{O}(2^L) = L2^{L-1} - k + 1 + \mathcal{O}(2^L).$$

When we consider the asymptotic frequency of occurrences, we note that $n \to \infty$ implies that $L \to \infty$. Therefore the asymptotic frequency of this type of occurrences is

$$\lim_{L\to\infty} \frac{k(2^L - 1) + \mathcal{O}(2^L)}{L2^{L-1} - k + 1 + \mathcal{O}(2^L)} = 0.$$

We conclude that the frequency of divided occurrences tends to zero. $\square$

Theorem 4.13. *The number $\mathcal{K}$, defined in* (4.10)*, is Minkowski normal.*

Proof of Theorem 4.13. Let $d = d_1, d_2, \cdots, d_k$ be an arbitrary block of length k, $d_i \in \mathbb{N}$. In order to determine the frequency of d in $\mathcal{K}$ it is sufficient to count the following binary blocks in $\mathcal{C}_2$.

- The frequency of
$$10^{d_k-1} \cdots 10^{d_2-1}10^{d_1-1}1, \tag{4.19}$$
- and the frequency of
$$10^{d_k-1} \cdots 10^{d_2-1}10^{d_1-1}0. \tag{4.20}$$

We argue this by considering the four different types of occurrences.

Firstly, it follows from Lemma 4.12 that the frequency of divided occurrences of d tends to 0. Furthermore, from Lemma 4.9, we find that the subpath that is associated to forming the block d is associated to the binary code

$$10^{d_k-1} \cdots 10^{d_2-1}10^{d_1-1}. \tag{A}$$

This corresponds to d occurring at the start of a continued fraction expansion.

The binary code associated to occurrences of d in the middle of a continued fraction expansion also results from Lemma 4.9, see equation (4.17). The difference with (A) is that another right move is needed in the Kepler tree, which preserves the block forever and causes it to occur in the middle. Therefore, the binary code associated to this type of occurrence is the same as that in (A) with a 1 appended. Hence

$$10^{d_k-1} \cdots 10^{d_2-1}10^{d_1-1}1. \tag{B}$$

Furthermore, d can also occur at the end of a continued fraction. Due to the fact that the Kepler rule alters the start of continued fraction expansions, these type of occurrences are descendants from the rational $[d_1, d_2, \cdots, d_k]$. In order to preserve the block d forever, another right move is needed. Using this and Lemma 4.8 we find that the corresponding binary code is

$$0^{d_k-2} \cdots 10^{d_2-1}10^{d_1-1}1, \tag{C}$$

where the last 1 results from the extra right move. However, occurrences of this binary code in $\mathcal{C}_2$ do not always correspond to an occurrence of d in $\mathcal{K}$. This is due to the fact that the digit 2 is used to form d_k. That is, d_k is formed from the digit 2, whereas in the other type of occurrences, the block d is formed from scratch. Hence for the binary code in (C) to correspond to an occurrence of d in $\mathcal{K}$, this occurrence of d should originate from a rational of the form $[2, b_2, \cdots, b_{j-1}, b_j]$. By Lemma 4.8, this corresponds to rationals that have a binary code given by

$$0^{b_j-2} \cdots 10^{b_2-1}10.$$

In other words, for (C) to correspond to an occurrence of d in $\mathcal{K}$, we need to consider occurrences of d that originate from rationals whose corresponding binary code ends in 10. If d is formed through a subpath that starts from such a rational, the binary code that is associated to this subpath is appended to that of the rational it originates from. From this and Lemma 4.9, we conclude that we can count these occurrences by looking at the frequency of the block

$$100^{d_k-2}\cdots 10^{d_2-1}10^{d_1-1}1 = 10^{d_k-1}\cdots 10^{d_2-1}10^{d_1-1}1. \tag{C*}$$

This is similar to (B). Moreover by counting the blocks in (A), we count (B) and (C*) as well. In order to prevent double counts, we append a 0 to the code in (A). In conclusion, in order to find the frequency of d in $\mathcal{K}$, it is sufficient to consider the asymptotic frequencies in $\mathcal{C}_2$ of (4.19) and (4.20). Both blocks occur with relative frequency

$$2^{-(d_1+\cdots+d_k+1)}.$$

This results from the fact that the binary codes are of length $d_1 + d_2 + \cdots + d_k + 1$ and that $\mathcal{C}_2$ is normal in base 2. Adding these frequencies gives the desired result

$$\frac{1}{2^{d_1+\cdots+d_k+1}} + \frac{1}{2^{d_1+\cdots+d_k+1}} = 2^{-(d_1+\cdots+d_k)}.$$

We conclude that $\mathcal{K}$ is Minkowski normal. □

The key idea in the proof is the unique correspondence between binary codes and continued fractions. Although the arguments in the proof refer to the Kepler tree, it is the coding that allows us to obtain frequencies. Moreover, the frequencies in (4.19) and (4.20) can be used to obtain frequencies in more general cases. This is due to the underlying structure that causes the normality. We elaborate on this generalisation after we have discussed the numerical experiment.

4.6.0 *Numerical verification and final remarks*

In addition to proving Minkowski normality of the number $\mathcal{K}$, we provide a numerical experiment that supports the proof from an empirical perspective. We briefly explain the setup of the experiment, argue its validity and show the result. The code that is used for this experiment can be found in the Appendix A.

In order to support the proof from an empirical perspective, we have plotted the distribution of the orbit of $\mathcal{K}$. First, we generate the Kepler tree up to the 12th level, see Listing A.1. In order to plot the distribution of the orbit $(\mathcal{T}^i\mathcal{K})_{i\geq 0}$, we approximate $\mathcal{T}^n\mathcal{K}$ by its 20-th convergent. That is, if $\mathcal{K} = [a_1, a_2, a_3, \cdots]$, we approximate $\mathcal{T}^n\mathcal{K} = [a_{n+1}, a_{n+2}, a_{n+3}, \cdots]$ by

$$\omega_{20}(\mathcal{T}^n\mathcal{K}) = [a_{n+1}, a_{n+2}, \cdots, a_{n+21}]$$

and use a 25 digit precision when doing so, see Listing A.2. Consequently, we approximate the distribution of the orbit through a histogram plot that uses bins of length 0.005. This resulting plot is featured in Figure 4.5. Comparing this to Figure 2.2, we see that they clearly resemble each other.

We argue the validity of this numerical experiment using ergodic theory. The main argument for this is based on the fact that the Gauss map is measure preserving with respect to the Minkowski question mark measure. More specifically, we use the fact that $([0, 1), \mathcal{B}, ?, \mathcal{T})$ is a measure preserving system. It then follows that, for any measurable function $f : [0, 1) \to \mathbb{R}$ and any $x \in [0, 1)$, the sequence $(f(\mathcal{T}^i x))_{i \geq 0}$ is stationary. Letting $f(x) = x$ and choosing $x = \mathcal{K}$, we find that the distribution of the constructed number $\mathcal{K}$ is stationary under iterations of $\mathcal{T}$. Therefore, plotting the first N iterations of the orbit can be interpreted as plotting N random variables that have the same distribution as the partial quotients of $\mathcal{K}$. Hence, by making a distribution plot of these N random variables, we implicitly plot the distribution of $\mathcal{K}$.

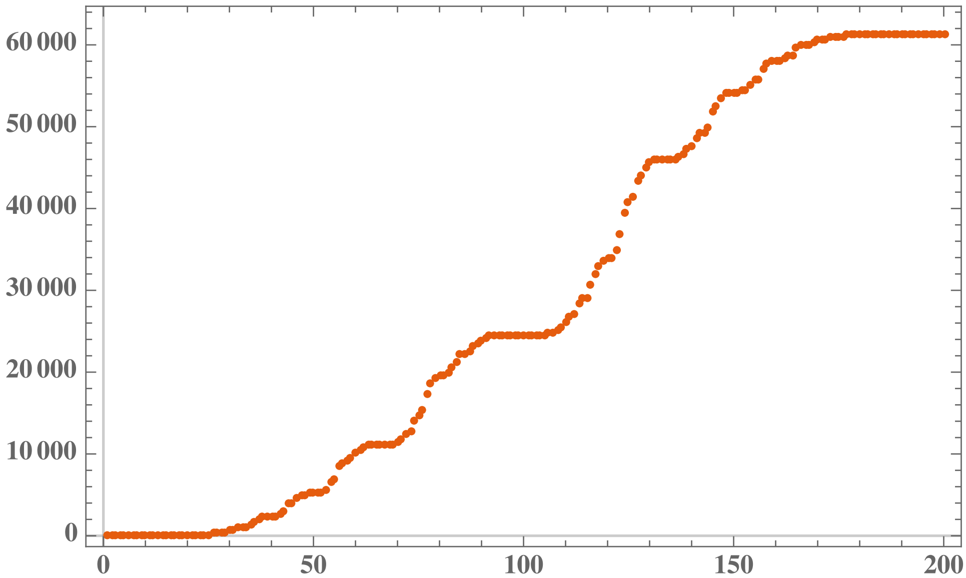


Figure 4.5: Plotting the orbit of $\mathcal{K}$ under $\mathcal{T}$.

CHAPTER 5

Other results and further research

In this chapter, we reflect on the construction and extend our results. We discuss the underlying structure of the Kepler tree to show that we can extend normality of the constructed number $\mathcal{K}$ to more general cases. One of the extensions shows that we can make local permutations in $\mathcal{K}$ to obtain a class of Minkowski normal numbers. Locality will be defined on the structure that underlies the construction that is done in the previous section. Furthermore, we show that the number that is obtained by concatenating the continued fraction expansions of the rationals in the Farey tree top-down, left-right, is such a permutation of $\mathcal{K}$. Therefore, we conclude that we can construct another concrete Minkowski normal number using the Farey tree. We close the chapter by suggesting topics for further research.

5.1.0 *Extending Minkowski normality of $\mathcal{K}$*

When constructing a normal number, it is the ordering that is chosen that determines the distribution. Apparently, ordering the rationals based on their denominator leads to the distribution given by the Gauss measure, e.g. see Vandehey [35]. Although the sequence of rationals in (3.11) is distributed according to the Lebesgue measure and not the Gauss, it is not that surprising that the number constructed by Adler, Keane and Smorodinsky is continued fraction normal. When we consider the frequency of occurrences of an arbitrary block $d = d_1, d_2, \cdots, d_k$ starting at the n-th position of a continued fraction expansion of a number in a uniformly distributed sequence, this frequency is given by the Lebesgue measure of the set $\mathcal{T}^{-n}\Delta(d)$ [1]. Gauss showed that the Perron-Frobenius operator of the $\mathcal{T}$ under λ is given by $\frac{1}{\log 2} \cdot \frac{1}{1+x}$. In other words, Gauss showed that, as $n \to \infty$, $\lambda(\mathcal{T}^{-n}\Delta(d))$ converges weakly to $\gamma(\Delta(d))$. Similarly, when we consider the sequence of rationals that is used in the construction, it should not be surprising that the number $\mathcal{K}$ is Minkowski normal. Namely, the sequence of rationals that is obtained by ordering the rationals in the Kepler tree top-down left-right, see (4.9), is distributed according to the Minkowski question mark. Then it follows that the frequency of occurrences of d, starting at the n-th position of a continued fraction expansion of a number in a Minkowski question mark distributed sequence, is given by the Minkowski measure of $\mathcal{T}^{-n}\Delta(d)$. As $\mu_?$ is $\mathcal{T}$-invariant, this measure is simply $\mu_?(\Delta(d))$. The fact that the sequence in (4.9) is distributed according to $\mu_?$ has implicitly been proved by Viader, Paradís and Bibiloni [37]. In the article, they first define a one-to-one correspondence $q : \mathbb{N} \to (0, 1)$. The first few terms of q are

$$
\begin{aligned}
q(1) &= [2] = 1/2 & \qquad q(5) &= [1,3] = 3/4 \\
q(2) &= [3] = 1/3 & q(6) &= [2,2] = 2/5 \\
q(3) &= [1,2] = 2/3 & q(7) &= [1,1,2] = 3/5 \\
q(4) &= [4] = 1/4 & q(8) &= [5] = 1/5,
\end{aligned}
$$

which result from the following definition. If $n = 2^{a_1} + 2^{a_2} + \cdots + 2^{a_k}$ with $0 \leq a_1 < a_2 < \cdots < a_k$, then:

$$q(n) := \begin{cases} [k+2] & \text{if } n = 2^k, \\ [a_1+1, a_2-a_1, a_3-a_2, \cdots, a_k - a_{k-1}+1] & \text{otherwise.} \end{cases} \tag{5.1}$$

Among other things, Viader, Paradís and Bibiloni prove that the sequence $(q(i))_{i]\geq 1}$ is distributed according to the Minkowski question mark measure.

Theorem 5.1. *[37, Theorem 2.7] For any $x \in [0,1]$, we have that*

$$\lim_{n\to\infty} \frac{\#\{q(i) \leq x : 1 \leq i \leq n\}}{n} = ?(x),$$

where $\#A$ denotes the cardinality of the set A.

We next show that the sequence of rationals in (4.9) is distributed according to the Minkowski question mark. More specifically, we prove that this sequence coincides with the sequence $(q(i))_{i\geq 1}$. Recall that the sequence in (4.9) is also represented by $([k_i])_{i\geq 1}$, where k_i denotes the block of digits that form the continued fraction expansion of the i-th rational in (4.9).

Corollary 5.2. The sequence $([k_i])_{i\geq 1}$ is distributed according to the Minkowski question mark measure. That is, for any $x \in [0,1]$, we have that

$$\lim_{n\to\infty} \frac{\#\{[k_i] \leq x : 1 \leq i \leq n\}}{n} = ?(x),$$

where $\#A$ denotes the cardinality of the set A.

Proof. We prove that $q(n) = [k_n]$ for all $n \in N$. It is clear that $q(1) = [k_1] = 1/2$. We next show that the Kepler rule coincides with:

$$\begin{tikzcd}[column sep=small] & q(n) \arrow[dl, dash] \arrow[dr, dash] & \\ q(2n) & & q(2n+1) \end{tikzcd},$$

which concludes the proof. Let $n = 2^{a_1} + 2^{a_2} + \cdots + 2^{a_k}$ with $0 \leq a_1 < a_2 < \cdots < a_k$. Suppose that $n = 2^l$ for some l. Then $2n = 2^{l+1}$ and $2n+1 = 2^0 +^{l+1}$. Using (5.1), we find

$$\begin{tikzcd}[column sep=small] & q(n) = [l+2] \arrow[dl, dash] \arrow[dr, dash] & \\ q(2n) = [(l+1)+2] = [(l+2)+1] & & q(2n+1) = [0+1, (l+1)-0+1] = [1, l+2] \end{tikzcd}.$$

Next, assume that $n = 2^{a_1} + 2^{a_2} + \cdots + 2^{a_k} \neq 2^l$. Then $q(n) = [a_1 + 1, a_2 - a_1, a_3 - a_2, \cdots, a_k - a_{k-1} + 1]$, and

$$\begin{aligned} 2n &= 2^{a_1+1} + 2^{a_2+1} + \cdots + 2^{a_k+1}; \\ 2n+1 &= 2^0 + 2^{a_1+1} + 2^{a_2+1} + \cdots + 2^{a_k+1}. \end{aligned}$$

Applying (5.1) to the above, we get

$$\begin{aligned} q(2n) &= [(a_1+1)+1, (a_2+1)-(a_1+1), (a_3+1)-(a_2+1), \cdots, (a_k+1) \\ &\quad - (a_{k-1}+1)+1] \\ &= [(a_1+1)+1, a_2-a_1, a_3-a_2, \cdots, a_k-a_{k-1}+1]; \\ q(2n+1) &= [0+1, (a_1+1)-0, (a_2+1)-(a_1+1), (a_3+1) \\ &\quad - (a_2+1), \cdots, (a_k+1)-(a_{k-1}+1)+1] \\ &= [1, (a_1+1), a_2-a_1, a_3-a_2, \cdots, a_k-a_{k-1}+1]. \end{aligned}$$

Hence our claim is true. Therefore $(q(i))_{i \geq 1}$ coincides with the sequence of rationals in (4.9) and Theorem 5.1 and Corollary 5.2 are equivalent. □

From this perspective, it is not that surprising that the constructed number $\mathcal{K}$ is Minkowski normal. Similarly, as the rationals in the even Stern-Brocot sequence are distributed according to the Minkowski question mark, see (4.4) and (4.8), it should not be surprising that we can construct a Minkowski normal number using this sequence. First, we discuss the structure that underlies the ordering of rationals and causes the normality. Then, we show that this structure can be used to construct a class of Minkowski normal numbers and provide an example using the Farey tree.

The continued fraction normality of x_{aks} results from the ordering of rationals based on their denominator. This ordering causes the sequence of rationals in (3.11) to be distributed uniformly and hence x_{aks} to be continued fraction normal. Minkowski normality of $\mathcal{K}$, however, results from a completely different underlying structure. The underlying structure in this case comes from fact that the rationals are ordered increasingly, based on the sum of the digits of their continued fraction expansion. That is, the l-th level of the Kepler tree contains all possible rationals that have a continued fraction expansion whose sum of digits is equal to $l + 2$. By ordering these top-down, left-right, the ordering is done as claimed. To see that the Kepler tree has this structure, we start by considering the root. The root of the tree, which corresponds to level 0, is given by $1/2 = [2]$. Then, every next level, the sum of digits of the continued fraction expansion is increased by 1 through the Kepler rule, see Lemma 4.4. Furthermore, the l-th level of the Kepler tree contains 2^l rationals, which is exactly the number of distinct[1] rationals that have a continued fraction expansion whose digits sum up to $l + 2$.

[1]We say that two rationals p/q and r/s are distinct if and only if $ps \neq qr$.

Proposition 5.3. There exist exactly 2^l distinct rationals that have a continued fraction expansion of which the sum of the digits equals $l+2$, $l \geq 0$. That is,

$$\#\left\{\frac{p}{q} \in [0,1) : \frac{p}{q} = [a_1, a_2, \cdots, a_n], \sum_{i=1}^{n} a_i = l+2\right\} = 2^l,$$

where $\#A$ denotes the cardinality of the set A.

Proof. This follows from [28, Problem 21]. Here, they state the following.

> "It is possible to write the positive integer n in $2^{n-1}-1$ ways as a sum of smaller positive integers. Two sums that differ in the order of terms only are now regarded as different. E.g. only the seven following sums add up to 4:
>
> $$\begin{array}{cccc} 1+1+1+1, & 1+1+2, & 2+2, & 1+3, \\ & 1+2+1, & & 3+1, \\ & 2+1+1." & & \end{array}$$

For continued fractions, we have to correct for the case where the last digit of the continued fraction expansion is equal to 1. Therefore, the number of allowed partitions of $l+2$, $l \geq 0$, is given by:

$$(2^{(l+2)-1}-1)-(2^{(l+1)-1}-1) = 2^{l+1}-2^l = 2^l.$$

Hence we conclude that there exist exactly 2^l distinct rationals that have a continued fraction expansion whose digits sum up to $l+2$. □

Due to this proposition, we conclude that $\mathcal{K}$ is a concrete example of a number that is obtained by concatenating all possible continued fraction expansions in increasing order, based on the sum of their digits. Next, we show that all such constructions are Minkowski normal. In order to prove this, we first introduce a result from Shiokawa and Uchiyama that proves normality of the locally permuted dyadic Champernowne number.

Lemma 5.4. *[31, Lemma 4] For any $l \geq 1$ let $c_{l,j}$, $1 \leq j \leq 2^l$, denote the possible binary blocks of length l. Arbitrarily subdivide each block $c_{l,j}$ into at most $h(l)$ parts, $1 \leq h(l) \leq l$, subject to the condition that every part should consist of only one or more consecutive digits in the original block. Let c'_l denote the sequence of length $l2^l$ obtained by concatenating the (at most) $h(l)2^l$ resulting subblocks in arbitrary order. If*

$$h(l) = o(l) \quad \text{as } l \to \infty.$$

Then the number $C' = 0.c'_1c'_2c'_3\cdots$ is normal in base 2.

Intuitively, we can interpret the lemma as follows. If we take the dyadic Champernowne number $\mathcal{C}_2$ and make local permutations, then the permuted number is also normal in base 2. Here, local means that the permutations takes place within the parts of the sequence that concatenates blocks of the same length. Furthermore, the condition related to the asymptotics of $h(l)$ implicitly states that the length of the resulting subblocks should be small relative to the length of the original block. It should be clear that if $h(l) = 1$ for all l, the conditions of the lemma are satisfied. This was also proved by Denker and Krämer [14], who implicitly stated the following.

Corollary 5.5. Let $\mathcal{C}_2$ be denoted by

$$\mathcal{C}_2 = 0.\, c_1^1\, c_2^1\, c_1^2\, c_2^2\, c_3^2\, c_4^2\, c_1^3\, c_2^3\, c_3^3\, c_4^3\, \cdots,$$

where $c_1^l c_2^l \cdots c_{2^l}^l$ denotes the concatenation of all binary codes in the l-th level of the binary Kepler tree, ordered from left to right. For all $l \in \mathbb{N}$, let π_l be a permutation of $\{1, 2, \ldots, 2^l\}$. Then

$$\mathcal{C}_2^\pi := 0.\, c_{\pi_1(1)}^1\, c_{\pi_1(2)}^1\, c_{\pi_2(1)}^2\, c_{\pi_2(2)}^2\, c_{\pi_2(3)}^2\, c_{\pi_2(4)}^2\, c_{\pi_3(1)}^3\, c_{\pi_3(2)}^3\, c_{\pi_3(3)}^3\, c_{\pi_3(4)}^3\, \cdots$$

is normal in base 2.

Due to the structure that underlies our construction, we can use this corollary to extend our results. Again, the key idea is the unique correspondence between binary codes of length l and continued fractions whose digits sum up to $l+2$. Let $[a_1, a_2, \cdots, a_n]$ be such that $\sum_{i=1}^n a_i = l + 2$, then recall that this correspondence is given by

$$[a_1, a_2, \cdots, a_n] \overset{\text{binary code}}{\longleftrightarrow} \underbrace{0^{a_n-2} \cdots 10^{a_2-1}10^{a_1-1}}_{\text{binary code of length l}}. \tag{5.2}$$

As Example 4.5 focusses on obtaining a binary code from a given continued fraction expansion, we provide the following example to show the correspondence from the opposite direction.

Example 5.6. Consider the binary code 1001010111 of length $l = 10$. If we rewrite this in the form that is given on the right-hand side (5.2), we find

$$0^{2-2}10^{3-1}10^{2-1}10^{2-1}10^{1-1}10^{1-1}10^{1-1}.$$

We conclude that the binary code 1001010111 corresponds (uniquely) to the continued fraction $[1, 1, 1, 2, 2, 3, 2]$. Note that the digits sum up to $l + 2 = 12$.

The proof of Theorem 4.13 shows that we can count arbitrary blocks in $\mathcal{K}$ through binary codes and explains why and how by referring to the structure of the Kepler tree. However, it is the coding that is important. Moreover, it is the explicit one-to-one correspondence between continued fraction expansions and binary codes that allows us to obtain frequencies and extend our results. This is due to the fact that divided occurrences are negligible and that the binary codes in (4.19) and (4.20) result from the coding that is used. That is, if we convert a continued fraction expansion $[a_1, a_2, \cdots, a_n]$ to its binary code $0^{a_n-2} \cdots 10^{a_2-1}10^{a_1-1}$, we can use the binary codes in (4.19) and (4.20) to obtain the frequency of $d = d_1, d_2, \cdots, d_k$ in $[a_1, a_2, \cdots, a_n]$. As such, we can extend the normality of $\mathcal{K}$ to more general cases.

Theorem 5.7. *Let the constructed number $\mathcal{K}$ be denoted by*

$$\mathcal{K} = [\kappa_1^1,\ \kappa_2^1,\ \kappa_1^2,\ \kappa_2^2,\ \kappa_3^2,\ \kappa_4^2,\ \kappa_1^3,\ \kappa_2^3,\ \kappa_3^3,\ \kappa_4^3,\ \cdots],$$

where $\kappa_1^l, \kappa_2^l, \cdots, \kappa_{2^l}^l$ is the concatenation of the continued fraction expansions of the rationals in the l-th level of the Kepler tree, ordered from left to right. Furthermore, for all $l \in \mathbb{N}$, let π_l be a permutation of $\{1, 2, \ldots, 2^l\}$. Then

$$\mathcal{K}^\pi := [\kappa^1_{\pi_1(1)},\ \kappa^1_{\pi_1(2)},\ \kappa^2_{\pi_2(1)},\ \kappa^2_{\pi_2(2)},\ \kappa^2_{\pi_2(3)},\ \kappa^2_{\pi_2(4)},\ \kappa^3_{\pi_3(1)},\ \kappa^3_{\pi_3(2)},\ \kappa^3_{\pi_3(3)},\ \kappa^3_{\pi_3(4)},\ \cdots]$$

is Minkowski normal.

Proof. The proof follows almost directly from Corollary 5.5. Let $d = d_1, d_2, \cdots, d_k$ be an arbitrary block of length k. Analogous to Lemma 4.12, we conclude that the asymptotic frequency of divided occurrences of d in $\mathcal{K}^\pi$ tends to zero. Subsequently, we note that $\mathcal{C}_2^\pi$ corresponds to the concatenation of the binary codes of the continued fraction expansions that are concatenated in $\mathcal{K}^\pi$. As these binary codes and continued fraction expansions are uniquely related by the correspondence in (5.2), we can count the number of occurrences of d in $\mathcal{K}^\pi$ by considering the frequency of (4.19) and (4.20) in $\mathcal{C}_2^\pi$. The rest of the proof then becomes analogous to the last part of the proof of Theorem 4.13. We conclude that $\mathcal{K}^\pi$ is Minkowski normal. □

In particular, Theorem 5.7 proves Minkowski normality of the number that is obtained by concatenating the continued fraction expansions of the rationals in the even Stern-Brocot sequence, see (4.4). This sequence corresponds to the level sets of the Farey tree, see the left subtree of Figure 4.1. Therefore, concatenating the continued fraction expansions of the rationals in the even Stern-Brocot sequence is equivalent with concatenating the continued fraction expansions of the rationals in the Farey tree top-down, left-right. The tree starts with $1/2 = [2]$ at the root and forms new rationals according to the tree rule displayed in Figure 5.1, see [7]. The ordering of the rationals that is obtained by this, is

$$\frac{1}{2},\quad \frac{1}{3}, \frac{2}{3},\quad \frac{1}{4}, \frac{2}{5}, \frac{3}{5}, \frac{3}{4},\quad \frac{1}{5}, \ldots .$$

It was shown by Kessebömer and Stratmann that this sequence is distributed according to $\mu_?$, see (4.8). Therefore it should not be surprising that the following holds.

Corollary 5.8. The number that is obtained by concatenating the continued fraction expansions of the rationals in the Farey tree top-down left-right is Minkowski normal.

Proof. It can be seen from the tree rules that, regardless of whether n is even or odd, the Farey tree rule increases the sum of the digits of the continued fraction expansion by 1 each next level. Therefore, the underlying structure of the tree is similar to that of the Kepler tree. Namely, the rationals are ordered increasingly, based on the sum of the digits of their continued fraction expansion. That is, the l-th level of the Farey tree contains all possible rationals that have a continued fraction expansion whose sum of digits is equal to $l + 2$. Hence by concatenating the continued fraction expansions of the rationals in the Farey tree top-down, left-right, we obtain a permutation of $\mathcal{K}$ that satisfies the conditions in Theorem 5.7. Hence we conclude that the number that is obtained by concatenating the continued fraction expansions of the rationals in the Farey tree top-down left-right is Minkowski normal. □

$$[a_1, a_2, \cdots, a_n]$$

$$[a_1, a_2, \cdots, (a_n + 1)] \quad [a_1, a_2, \cdots, (a_n - 1), 2]$$

(a)

$$[a_1, a_2, \cdots, a_n]$$

$$[a_1, a_2, \cdots, (a_n - 1), 2] \quad [a_1, a_2, \cdots, (a_n + 1)]$$

(b)

Figure 5.1: The rule of the Farey tree for (a) n is odd and (b) n is even.

The structure that is apparent in both the Kepler and Farey tree thus allows us to construct Minkowski normal numbers from these trees. In order to prove Minkowski normality in the case of the Farey tree, we exploit this underlying structure and extend the Minkowski normality of $\mathcal{K}$ to the number obtained from the Farey tree. This extension uses results of Shiokawa and Uchiyama, who extended normality of the dyadic Champernowne number. The extension in Theorem 5.7 however is based on a specific case of the extension of Shiokawa and Uchiyama. That is, we use $h(l) = 1$ for all $l \in \mathbb{N}$, see Lemma 5.4. By doing so, we preserve the underlying structure and hence - in some way - preserve normality. We have not been able to prove a full analogue of Shiokawa and Uchiyama's result, which is therefore a topic of further research. This and other topics of further research are briefly discusses in the next section.

5.2.0 *Further research*

One of the possible topics of further research is to develop an analogue of Shiokawa and Uchiyama's work. The full analogue of their work, would approximately be as follows.

Minkowski normality analogue of Lemma 5.4 *For any $l \geq 1$ let $\kappa_{l,j}, 1 \leq j \leq 2^l$, denote the possible continued fraction expansions whose digits sum up to $l + 2$. Arbitrarily subdivide each continued fraction expansion $\kappa_{l,j}$ into at most $h(l)$ parts, $1 \leq h(l) \leq l + 1$, subject to the condition that every part should consist of only one or more consecutive digits in the original continued fraction expansion. Let κ'_l denote the sequence of length $(l+2)2^{l-1}$ obtained from concatenating the (at most) $h(l)2^l$ resulting subblocks in arbitrary order. If*

$$h(l) = o(l) \quad \textit{as } l \to \infty.$$

Then the number $\mathcal{K}' = [\kappa'_1, \kappa'_2, \kappa'_3, \cdots]$ is Minkowski normal.

One of the reasons that we cannot extend normality to this general case, is that the we can no longer use the normality of $\mathcal{C}_2^\pi$ to count frequencies. That is, when we break up continued fraction expansions into smaller parts, one creates subblocks of which the sum of its digits will vary and the composition of binary codes will change. Consider for instance the continued fraction $[2, 1, 1, 3]$, which corresponds to the binary code 01110. Suppose we break this up into $[2]$ and $[1, 1, 3]$. Then these correspond to the binary codes $\varnothing$ and 011 respectively. Conversely, break up 01110 into the blocks 011 and 10. These binary codes correspond, respectively, to the continued fraction expansions $[1, 1, 3]$ and $[1, 3]$. This shows that the underlying structure is not preserved for $h(l) \neq 1$ and we cannot extend normality along similar lines. Hence further research could be aimed at finding a Minkowski analogue of Lemma 5.4. In particular, it is yet unknown if there exists an asymptotic condition that guarantees Minkowski normality of the permuted number and what this condition should look like.

Also, as observed by Belshaw, one problem with constructed numbers is that the expansion of a constructed number is prone to be deterministic instead of random. As such, Belshaw introduced the notion of strong normality. This concept has been introduced for normality in a base, but not for other types of normality [3]. Therefore, one could try to generalise this notion for other types of normality. Furthermore, one could consider finding an analogue of (ϵ, k)-normality for Minkowski normality. This would be somewhat similar to the work of Adler, Keane and Smorodinsky or that of Vandehey, who extended their work. That is, Adler, Keane and Smorodinsky introduced m-goodness to show that almost all rationals with denominator at most m have a continued fraction expansion whose digits have good small-scale normality properties and Vandehey was able to obtain asymptotics for this. In a similar way, one could consider the rationals that have a continued fraction expansion whose sum is at most N and try to obtain asymptotics on the number of rationals that have good small-scale normality properties. As we have seen, there is a strong relation between Minkowski normality and the ordering of rationals based on the sum of the digits of their continued fraction expansion. Looking at the number of rationals that have good small-scale properties, can therefore be approached by considering specific integer partitions. Regardless of the approach, it is expected that such extensions will make it possible to further generalise constructions of Minkowski normal numbers.

The previously described topics are related to extending our results. Other topics could also be related to developing similar results for different types of continued fractions or measures. A specific example is to consider the work of Boca and Linden [6], who have studied analogues of the Minkowski question mark measure that are related to continued fractions with even and odd partial quotients. Last but not least, up to this date, there is no known application for normal numbers. Due to the random nature, they are linked to random number generators. However, a true application remains to be discovered. Therefore, further research could also be aimed at finding applications for normal numbers.

Chapter 6

Conclusion

Different number representations correspond to different notions of normality. The decimal expansion, for instance, corresponds to normality in a base and is related to the Lebesgue measure. Similarly, the continued fraction expansion corresponds to a type of normality that comes from the Gauss measure, which is also related to the Lebesgue measure. From both the perspective of Gauss as well as that of Lebesgue, it follows that almost all numbers are normal and almost all numbers are continued fraction normal. In this thesis, however, we considered the continued fraction expansion and the Minkowski question mark measure. This measure is singular with respect to the Gauss and the Lebesgue measure. Therefore, we took a completely different view on numbers. We used this measure to introduce the notion of Minkowski normality, see Definition 4.1, which is another type of normality for the continued fraction expansion. It turns out that also from this perspective, almost all numbers are normal. That is, it turns out that $\mu_?$-almost all numbers are Minkowski normal, see Theorem 4.2.

More importantly, we have constructed a concrete Minkowski normal number $\mathcal{K}$, see (4.10), and used this to construct a class of Minkowski normal numbers. The initial construction considers the ordering of the rationals that is obtained from the Kepler tree. The key idea that we have used to prove normality, is that we can create a binary analogue by associating a binary code to each rational in the Kepler tree. This binary analogue resulted in the construction of the dyadic Champernowne number $\mathcal{C}_2$, which is known to be normal in base 2. Subsequently, we showed that we can identify binary codes that correspond to occurrences of arbitrary blocks in $\mathcal{K}$. Therefore, we were able to extend the base 2 normality of $\mathcal{C}_2$ to Minkowski normality of $\mathcal{K}$. This forms the main result of the thesis, which is stated in Theorem 4.13. Furthermore, we have identified the underlying structure of the Kepler tree that causes Minkowski normality of the constructed number. That is, Minkowski normality results from the fact that the rationals are ordered increasingly, based on the sum of the digits of their continued fraction expansion. Based on this underlying structure, we have provided a generalised construction of a Minkowski normal number. This generalisation proves the existence of a class of Minkowski normal numbers, see Theorem 5.7. The theorem states that any construction that concatenates the continued fraction expansions of all rationals, ordered increasingly, based on the sum of the digits of their continued fraction expansion, results in a number that is Minkowski normal. Again, it was the explicit one-to-one correspondence between continued fraction expansions and binary codes that allowed us to prove normality.

The importance of the results in this thesis come from considering the Minkowski measure. Besides the fact that only three constructions have been developed for normality related to the continued fraction expansion, the corresponding type of normality has always been related to the Lebesgue measure. Minkowski normality, however, relies on a measure that is singular with respect to the Lebesgue measure. Therefore, it takes a completely different view on numbers. As almost all numbers are normal from both points of view, we conclude that normality is just a matter of perspective.

Acknowledgements

When considering the options for my research project, there was a clear double split: at university or a company and in Utrecht or abroad. However, I had no doubt about who to ask as my first supervisor: dr. Karma Dajani. She agreed to be my supervisor and, together, we discussed possible thesis topics. When she mentioned normal numbers and the Minkowski question mark, I had no idea what this encompassed but it had caught my interest. My decision became clear: I wanted to write my thesis at Utrecht University (under Karma's supervision) and - if possible - do part of the research abroad. With help from Karma, I got the opportunity to become a visiting scholar at the George Washington University, where I spent two and a half months under supervision of her former promotor: Prof. Robbie Robinson.

Apart from the aforementioned, Karma's door has always been open for discussions, (personal) questions and/or a cup of coffee. Hence, for these reasons and more, I want to express my sincerest gratitude to Karma. Furthermore, I want to thank the people at GWU for accepting me as a visiting scholar and being a part of my experience abroad. In particular, I want to thank Professor Robinson for supervising me while at GWU and giving me a hint of what it's like to do a PhD. Moreover, I want to thank Prof. dr. Sjoerd Verduyn Lunel for being my second reader and providing me with feedback.

Furthermore, I want to acknowledge Tristan van Leeuwen and Roberto Fenández for writing recommendation letters to gain financial support for my time in Washington D.C. Moreover, I want to thank the A.F. Monna Fund, De Fundatie van de Vrijvrouwe van Renswoude te Delft and the George Washington University for supporting me financially.

Last but not least, I want to thank the people who are close to me. They have been indispensable in maintaining a balance between my thesis and life outside of mathematics. Thank you all.

Appendix A

Mathematica code

Listing A.1: Generating the Kepler tree

```
left[list_] := Prepend[Drop[list, 1], list[[1]] + 1]
right[list_] := Prepend[list, 1]
brancha[list_] := {left[list], right[list]}
branchl[level_] := Flatten[Map[brancha, level], 1]
tree = NestList[branchl, {{1, 1}}, 12]
ftree = Flatten[tree]
```

Listing A.2: Determining successive convergents

```
FCF[list_] := FromContinuedFraction[Prepend[list, 0]]
sftree = Table[Take[ftree, {k, k + 20}], {k, 1, Length[ftree]
        - 21}];
rats = Map[FCF, sftree]
nrats = N[rats, 25]
```

Listing A.3: Plotting the convergents

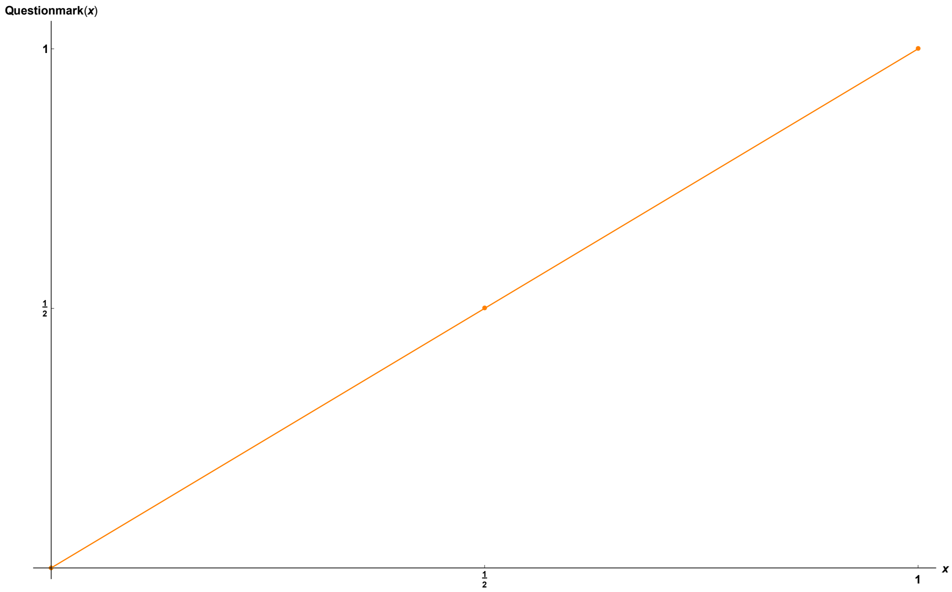
```
ListPlot[Accumulate[BinCounts[nrats, {0, 1, .005}]]]
```

Listing A.4: Constructing the Minkowski ?(·)

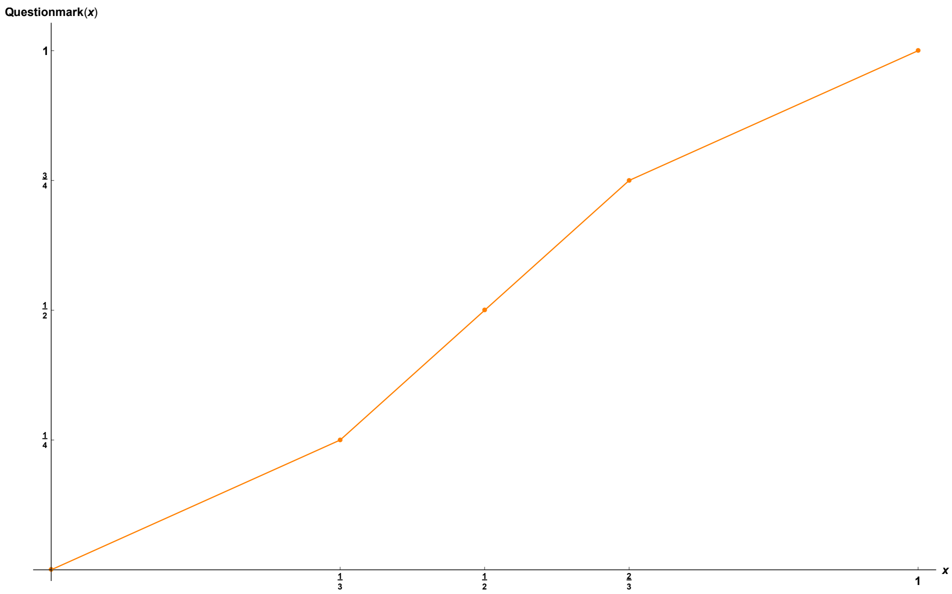
```
n = 1;
For[i = 1, i < n + 1, i ++, print[" i=", i, " list=", list];
 For[templist = list; j = 2, j < Length[list] + 1 , j++,

     Append[templist, (Numerator[Part[list, j - 1]] +
          Numerator[Part[list, j]])/(Denominator[Part[list,
                j - 1]] +
          Denominator[Part[list, j]])];
   ];
  list = Sort[templist];
  ]

myD[n_] := Table[k 2^(-n), {k, 0, 2^n}]

ListLinePlot[Transpose[{list, myD[n]}],
 AxesLabel -> {x, Questionmark[x]},
 AxesStyle -> Directive[Black, 14],
 PlotMarkers -> [Automatic, 10],
 PlotStyle -> {Orange},Ticks ->{list, myD[n]}]
```

Appendix B

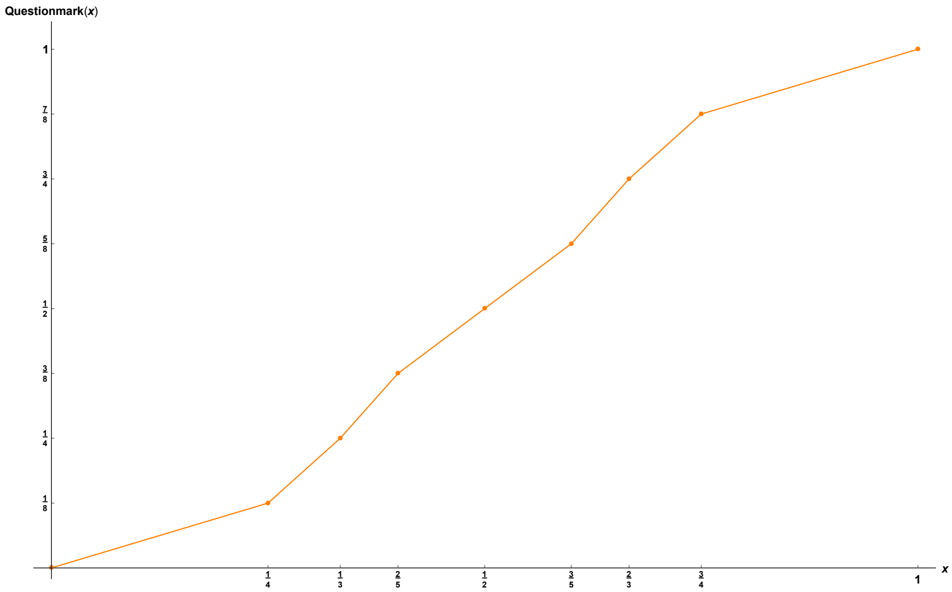
Approximations of the Minkowski question mark



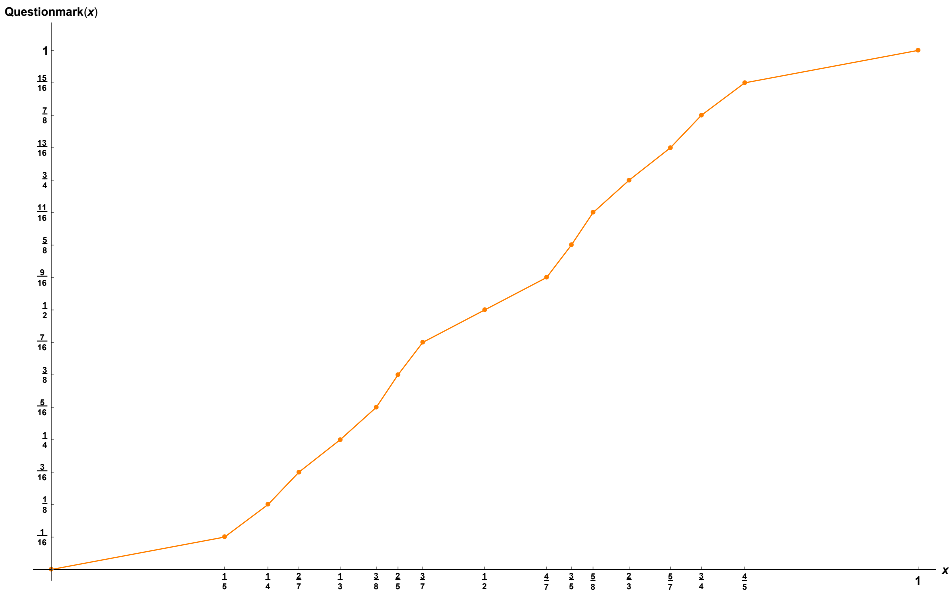
(a) Approximation of order $n = 1$



(b) Approximation of order $n = 2$



(c) Approximation of order $n = 3$



(d) Approximation of order $n = 4$

Bibliography

[1] Adler, R., Keane, M., and Smorodinsky, M. (1981). A construction of a normal number for the continued fraction transformation. *Journal of Number Theory*, 13(1):95–105.

[2] Becher, V. and Yuhjtman, S. A. (2017). On absolutely normal and continued fraction normal numbers. *arXiv preprint arXiv:1704.03622*.

[3] Belshaw, A. (2005). *On the normality of normal numbers*. PhD thesis, Simon Fraser University.

[4] Belshaw, A. and Borwein, P. (2013). Champernowne's Number, Strong Normality, and the X Chromosome. In *Springer Proceedings in Mathematics and Statistics*, pages 29–44. Springer, New York, NY.

[5] Besicovitch, A. S. (1935). The asymptotic distribution of the numerals in the decimal representation of the squares of the natural numbers. *Mathematische Zeitschrift*, 39(1):146–156.

[6] Boca, F. P. and Linden, C. (2018). On Minkowski type question mark functions associated with even or odd continued fractions. *Monatshefte für Mathematik*, 187(1):35–57.

[7] Bonanno, C. and Isola, S. (2008). Orderings of the rationals and dynamical systems. *arXiv preprint arXiv:0805.2178*.

[8] Catt, E., Coons, M., and Velich, J. (2016). Strong normality and generalized Copeland-Erdos numbers. *Integers*, 16:A11.

[9] Champernowne, D. G. (1933). The Construction of Decimals Normal in the Scale of Ten. *Journal of the London Mathematical Society*, s1-8(4):254–260.

[10] Copeland, A. H. and Erdös, P. (1946). Note on normal numbers. *Bulletin of the American Mathematical Society*, 52(10):857–861.

[11] Dajani, K. and Dirksen, S. (2018). Introduction to Ergodic Theory. *UU Course Notes*.

[12] Davenport, H. and Erdős, P. (1952). Note on normal decimals. *Canadian Journal of Mathematics*, 4(4):58–63.

[13] Denjoy, A. (1938). Sur une fonction réelle de Minkowski. *J. Math. Pures Appl*, 17(4):105–151.

[14] Denker, M. and Krämer, K. F. (1992). Upper and lower class results for subsequences of the Champernowne number. In *Ergodic theory and related topics III, Springer*, chapter 6, pages 83–89. Springer.

[15] Dushistova, A. A. and Moshchevitin, N. G. (2012). On the derivative of the Minkowski question mark function ?(x). *Journal of Mathematical Sciences*, 182(4):463–471.

[16] Émile Borel, M. (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27(1):247–271.

[17] Faber, G. (1910). Über stetige Funktionen. *Mathematische Annalen*, 69(3):372–443.

[18] Girgensohn, R. (1996). Constructing Singular Functions via Farey Fractions. *Journal of Mathematical Analysis and Applications*, 203(1):127–141.

[19] Iosifescu, M. and Kraaikamp, C. (2002). *Metrical Theory of Continued Fractions*, volume 547. Springer Netherlands, Dordrecht.

[20] Kamae, T. and Keane, M. (1997). A simple proof of the ratio Ergodic theorem. *Osaka Journal of Mathematics*, 34(3):653–657.

[21] Kepler, J. (1997). *The Harmony of the World.* American Philosophical Society, ilustrated edition.

[22] Kesseböhmer, M. and Stratmann, B. O. (2008). Fractal analysis for sets of non-differentiability of Minkowski's question mark function. *Journal of Number Theory*, 128(9):2663–2686.

[23] Kesseböhmer, M. and Stratmann, B. O. (2010). A dichotomy between uniform distributions of the Stern-Brocot and the Farey sequence. *arXiv preprint arXiv:1009.1823.*

[24] Madritsch, M. G. and Mance, B. (2016). Construction of μ -normal sequences. *Monatshefte für Mathematik*, 179(2):259–280.

[25] Madritsch, M. G., Scheerer, A.-M., and Tichy, R. F. (2018). Computable absolutely Pisot normal numbers. *Acta Arithmetica*, 184(1):7–29.

[26] Nakai, Y., Shiokawa, I., and Wagner, G. (1992). Discrepancy estimates for a class of normal numbers. *Acta Arithmetica*, 62(3):271–284.

[27] Niven, I. (2014). *Irrational Numbers.* The Mathematical Association of America.

[28] Pólya, G. and Szegö, G. (1972). *Problems and Theorems in Analysis I.* Springer-Verlag Berlin Heidelberg.

[29] Postnikov, A. G. (1960). Arithmetic modeling of random processes. *Trudy Matematicheskogo Instituta imeni VA Steklova*, 57:3–84.

[30] Salem, R. (1943). On Some Singular Monotonic Functions Which Are Strictly Increasing. *Transactions of the American Mathematical Society*, 53(3):427–439.

[31] Shiokawa, I. and Uchiyama, S. (1975). On some properties of the dyadic Champernowne numbers. *Acta Mathematica Academiae Scientiarum Hungaricae*, 26(1-2):9–27.

[32] Spreij, P. (2016). Measure Theoretic Probability. *UvA Course Notes.*

[33] Vandehey, J. (2016a). Absolutely abnormal and continued fraction normal numbers. *Bulletin of the Australian Mathematical Society*, 94(02):217–223.

[34] Vandehey, J. (2016b). New normality constructions for continued fraction expansions. *Journal of Number Theory*, 166:424–451.

[35] Vandehey, J. (2017). Non-trivial matrix actions preserve normality for continued fractions. *Compositio Mathematica*, 153(2):274–293.

[36] Vepstas, L. (2008). On the Minkowski measure. *arXiv preprint arXiv:0810.1265.*

[37] Viader, P., Paradís, J., and Bibiloni, L. (1998). A New Light on Minkowski's ?(x) Function. *Journal of Number Theory*, 73(2):212–227.